2011

Feet on the potential energy surface, head in the π clouds

Quentin Anthony Smith
Iowa State University

Feet on the potential energy surface, head in the π clouds

by

Quentin Anthony Smith

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Physical Chemistry

Program of Study Committee:
Mark S. Gordon, Major Professor
Theresa Windus
William Jenks
Jim Evans
Pat Thiel

Iowa State University
Ames, Iowa
2011

This work is dedicated, in equal measure, to electrons and to coffee.
Without them, none of this would have been possible.

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

Overview

The landscape of a potential energy surface is marked by chemically interesting features. Hills and valleys correspond to transition states and reactive intermediates; the deepest valley gives the most stable configuration. Mapping these features for individual molecules and for the interactions between molecules is one of the goals of computational chemistry.

The dispersion energy is a weak attractive force in intermolecular interactions. Dispersion energy results from a purely quantum mechanical effect, in which instantaneous multipoles on one molecule induce multipoles on another. Among neutral atoms or molecules that lack permanent multipole moments, the dispersion interaction is the principal attractive force. Dispersion also plays a significant role in the interaction between molecules with diffuse π clouds. This interaction is often difficult to capture with standard computational chemistry methods, so a comparison of the results obtained with various methods is itself important.

This work presents explorations of the potential energy surface of clusters of atoms and of the interactions between molecules. First, structures of small aluminum clusters are examined and classified as ground states, transition states, or higher-order saddle points. Subsequently, the focus shifts to dispersion-dominated π-π interactions when the potential energy surfaces of benzene, substituted benzene, and pyridine dimers are explored. Because DNA nucleotide bases can be thought of as substituted heterocycles, a natural extension of the substituted benzene and pyridine investigations is to model paired nucleotide bases. Finally, the success of the dispersion studies inspires the development of an extension to the computational method used, which will enable the dispersion energy to be modeled – and the potential energy surface explored – in additional chemical systems.

Dissertation Organization

This introduction describes the effective fragment potential (EFP) method and the various quantum mechanical methods used in subsequent chapters of this dissertation. The next chapter details an *ab inito* quantum mechanical study of 13-atom aluminum clusters, while the following three chapters concern EFP studies of aromatic dimers in which dispersion energy makes a significant contribution to the attraction between monomers. The final chapter prior to the Conclusions describes theory and code development toward a means of computing dispersion energy in mixed *ab inito*-EFP systems.

Chapter 2 details a study of the structures and energies of the small aluminum clusters Al_{13} and Al_{13}^-, involving some of the highest-level *ab initio* computations performed on these clusters to date. Most of the previous theoretical studies on these clusters used density functional methods that did not always converge to the correct minimum-energy structures. The electron affinity of Al_{13} was calculated and found to agree extremely closely with the experimental value, giving credence to the *ab inito* results.

Chapters 3, 4, and 5 describe EFP studies of substituted benzene dimers, pyridine dimers, and DNA nucleotide base pairs, respectively. All of these systems involve aromatic rings with π electron clouds that can give rise to strong dispersion energy interactions between molecules. Because dispersion interactions arise from electron correlation, they are difficult to model with density functional theory or with computationally inexpensive, low-level *ab inito* methods that do not properly account for this effect. The EFP method is shown to model dispersion interactions with an accuracy approaching that of high-level *ab inito* methods, at a fraction of the computational cost.

Finally, Chapter 6 presents progress toward the development of a theoretical model to calculate the dispersion energy term in the interaction between a molecule modeled with EFP and a molecule modeled with an *ab initio* method, such as Hartree-Fock or second-order perturbation theory (MP2).

Theoretical Background

Ab inito methods

Ab inito, "from the beginning," refers to a set of quantum mechanical methods derived from the first principles of quantum mechanics. Generally, this description is applied to methods that make certain approximations in order to solve the time-dependent Schrödinger equation[1-5]

$$\frac{\partial \Psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_n;t)}{\partial t} = -\frac{i}{\hbar}\hat{H}\Psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_n;t) \tag{1}$$

The wavefunction, Ψ, is a state function in that it contains all the information about the atomic or molecular system. Ψ is a function of the position $\vec{x}_n$ of particle *n* (here, the particles are electrons and nuclei) and time t. $\hbar$ is Planck's constant, *h*, divided by 2π; *i* is the square root of -1; and $\hat{H}$ is the Hamiltonian operator, a function that accounts for the kinetic and potential energy of the system.

The first approximation made by many *ab inito* methods is to consider only the states in which the position and the time components of the wavefunction are separable. Then the wavefunction can be expressed as a spatial part multiplied by a time-dependent part:

$$\Psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_n;t) = \psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_n) f(t) \tag{2}$$

This allows the simplification of the Schrödinger equation in (1) to its time-independent form,

$$\hat{H}\psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_n) = E\psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_n) \tag{3}$$

where *E* is the total energy of the system. Typically, the lowest energy state is sought; solving the equation then means finding the optimal positions for every particle to give the lowest value of *E*.

For a chemical system, the Hamiltonian operator takes the form

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} \tag{4}$$

where the terms correspond to the kinetic energy of the nuclei, the kinetic energy of the electrons, the potential energy of the nuclear-electronic attraction, the potential energy of the electron-electron repulsion, and the potential energy of the nuclear-nuclear repulsion, respectively. These terms can be written explicitly as

$$\hat{H} = -\frac{1}{2}\sum_{A=1}^{n}\frac{\nabla^2}{m_A} - \frac{1}{2}\sum_{i=1}^{e}\nabla_i^2 - \sum_{i=1}^{e}\sum_{A=1}^{n}\frac{Z_A}{\vec{r}_{iA}} + \sum_{i=1}^{e}\sum_{j<i}^{e}\frac{1}{\vec{r}_{ij}} + \sum_{A=1}^{n}\sum_{B<A}^{n}\frac{Z_A Z_B}{\vec{r}_{AB}} \tag{5}$$

where ∇ is the Laplacian operator, m_A is the ratio of the mass of a nucleus to the mass of an electron, Z_A is the charge of nucleus A, $\vec{r}_{ij}$ is the distance between particle *i* and particle *j*, and the summations go over all of the nuclei (*n*) or electrons (*e*) in the system. Although Eq. (5) gives the true Hamiltonian, when solving the Schrödinger equation, the last two terms of Eq. (5) give rise to a "many-body problem" in which the position of each particle depends on the position of every other particle. Simplifications must be made to these terms to make the problem tractable.

The first simplification is the Born-Oppenheimer approximation[6], which assumes the wavefunction can be separated into a nuclear part and an electronic part:

$$\psi_{total}(\vec{x}_1^{nuc},\vec{x}_2^{nuc},...;\vec{x}_1^{elec},\vec{x}_2^{elec},...) = \psi_{nuc}(\vec{x}_1^{nuc},\vec{x}_2^{nuc},...)\psi_{elec}(\vec{x}_1^{elec},\vec{x}_2^{elec},...) \tag{6}$$

where $\vec{x}_{nuc}$ gives the coordinates of the nuclei and $\vec{x}_{elec}$ gives the coordinates of the electrons. In physical terms, Eq. (6) says that the nuclei are stationary with respect to the electrons, a reasonable approximation in many cases because the velocity of nuclei is much smaller than that of the much lighter electrons. This approximation eliminates the nuclear kinetic energy term ($\hat{T}_n$ in Eq. (4)) from the Hamiltonian, and it allows the nuclear potential energy to be calculated once and held constant for the duration while the Schrödinger equation is solved. The Hamiltonian of Eq. (5) is then reduced to

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{e}\nabla_i^2 - \sum_{i=1}^{e}\sum_{A}\frac{Z_A}{\vec{r}_{iA}} + \sum_{i=1}^{e}\sum_{j<i}^{e}\frac{1}{\vec{r}_{ij}} \tag{7}$$

Eq. (7) contains no nuclear energy terms and depends only on the charge of the stationary nuclei. When used as the Hamiltonian in the time-independent Schrödinger equation of Eq. (3), it gives the electronic Schrödinger equation:

$$\hat{H}_{elec}\psi_{elec} = E_{elec}\psi_{elec} \tag{8}$$

After the electronic Schrödinger equation is solved, the nuclear part can be solved separately. The nuclei are then permitted to change position within the field produced by the electrons, determined in Eq. (8). The total energy of the system is the sum of the electronic energy E_{elec} from Eq. (8) and the repulsive nuclear potential energy.

Further approximations are needed to treat the electron-electron interactions ($\hat{V}_{ee}$ in Eq. (4)) in systems with more than one electron. The most basic approximation is to replace explicit electron-electron interactions with an averaged interaction and solve self-consistently. Each electron is subjected not to every other individual electron, but to the mean field they create. This approximation is known as the self-consistent field (SCF) or Hartree-Fock[7-10] (HF) method.

The one-electron Fock operator for electron 1 is given by

$$\hat{F}_a(1) = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{n}\frac{Z_A}{\vec{r}_{1A}} + \sum_{j=1}^{e}[\hat{J}_j(1) - \hat{K}_j(1)] \tag{9}$$

where $\hat{J}$ is the Coulomb operator and $\hat{K}$ is the exchange operator. The wavefunction φ_a of an electron is a spin-orbital, which takes the form

$$\varphi_a(\vec{x}) = \varphi_a(\vec{r};\omega) = \begin{cases}\phi_a(\vec{r})\alpha(\omega)\\ \phi_a(\vec{r})\beta(\omega)\end{cases} \tag{10}$$

where $\phi_a(\vec{r})$ is the spatial component and the spin component is either α or β (spin up or spin down). For simplicity, the discussion may be restricted to closed shell chemical systems – those in which all electrons are paired in orbitals (paired electrons have wavefunctions with identical spatial components but opposite spins). In this case, the Fock operator for electron 1 becomes

$$\hat{F}_a(1) = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{n}\frac{Z_A}{\vec{r}_{1A}} + \sum_{j=1}^{e/2}[2\hat{J}_j(1) - \hat{K}_j(1)] \tag{11}$$

The one-electron operators $\hat{J}$ and $\hat{K}$ act on orbitals via the following equations:

$$\hat{J}_j(1)\phi_a(1) = \phi_a(1)\int d\vec{r}_2 \cdot \left|\phi_j(2)\right|^2 \vec{r}_{12}^{-1} \tag{12}$$

representing the average local potential at $\vec{r}_1$ arising from an electron in orbital ϕ_j, and

$$K_j(1)\phi_a(1) = \phi_j(1)\int d\vec{r}_2 \cdot \phi_j^*(2)\vec{r}_{12}^{-1}\phi_a(2) \tag{13}$$

The exchange energy is a purely quantum mechanical effect arising from symmetry considerations: the fact that a wavefunction for multiple indistinguishable fermions, like electrons, must change sign if the labels of two particles are interchanged.

$$\psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_i,...,\vec{x}_j,...,\vec{x}_n) = -\psi(\vec{x}_1,\vec{x}_2,...,\vec{x}_j,...,\vec{x}_i,...,\vec{x}_n) \tag{14}$$

The Fock operator acts on a set of orbitals $\{\phi_1,\ldots,\phi_n\}$ via

$$\hat{F}_i \phi_i = \varepsilon_i \phi_i \tag{15}$$

to give orbital energies $\varepsilon_1,\ldots,\varepsilon_n$.

In quantum chemical computations, each one-electron orbital itself is a linear combination of basis functions. The ith molecular orbital ϕ_i can be represented by a summation over basis functions $\{\chi_1,\ldots, \chi_n\}$:

$$\phi_i = \sum_{\nu=1}^{N} C_{\nu i}\chi_\nu \tag{16}$$

where $C_{\mu i}$ are LCAO[11] (linear combination of atomic orbitals) coefficients. Basis sets are usually comprised of Gaussian functions. An infinite number of basis functions in linear combination could be used to construct any other well-behaved function, such as a (physically meaningful) wavefunction. While using an infinite number of basis functions is not computationally tractable, using as large a basis set as possible gives more flexibility to the construction of orbitals.

The Hartree-Fock equations can be written in matrix form. Using the expression in Eq. (16) for the molecular orbitals and inserting into Eq. (15), the Fock operator acting on electron 1 can be expressed as

$$\hat{F}(1)\sum_\nu C_{\nu i}\chi_\nu(1) = \varepsilon_i \sum_\nu C_{\nu i}\chi_\nu(1) \tag{17}$$

Multiplying by $\chi_\mu^*(1)$ and integrating over electron 1 gives

$$\sum_\nu C_{\nu i}\int d\vec{r}_1 \chi_\mu^*(1)\hat{F}(1)\chi_\nu(1) = \varepsilon_i \sum_\nu C_{\nu i}\int d\vec{r}_1 \chi_\mu^*(1)\chi_\nu(1) \tag{18}$$

The integrals on the left are defined as matrix elements of the Fock operator, and those on the right as elements of the overlap matrix, S. Then,

$$\sum_\nu F_{\mu\nu}C_{\nu i} = \varepsilon_i \sum_\nu S_{\mu\nu}C_{\nu i} \tag{19}$$

where F is the Fock matrix, or

$$FC = SC\varepsilon \tag{20}$$

The energy matrix ε can be made diagonal via a unitary transformation, which only changes the wavefunction by a phase factor; this affects nothing observable. The Fock matrix was described previously as consisting of matrix elements of the one-electron Fock operator, given in Eq. (11). More specifically,

$$F_{\mu\nu} = \int d\vec{r}_1 \chi_\mu^*(1)\hat{F}(1)\chi_\nu(1) = \int d\vec{r}_1 \chi_\mu^*(1)\left[\hat{h}(1) + \sum_a 2\hat{J}_a(1) - \hat{K}_a(1)\right]\chi_\nu(1) \tag{21}$$

Using the notation $(rs|pq) = \int d\vec{r}_1 d\vec{r}_2 \chi_r^*(1)\chi_s(1)\vec{r}_{12}\chi_p^*(2)\chi_q(2)$, this simplifies to

$$F_{\mu\nu} = h_{\mu\nu} + \sum_a 2(\mu\nu|aa) - (\mu a|a\nu) = h_{\mu\nu} + \sum_{\rho\sigma}\left(\sum_a C_{\rho a}C_{a\sigma}\right)\left[2(\mu\nu|\rho\sigma) - (\mu\rho|\sigma\nu)\right] \tag{22}$$

The density matrix for the orbital (LCAO) coefficients, *D*, is defined as the product of the coefficient matrix and its transpose,

$$D = C^T C \tag{23}$$

So,

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\rho\sigma} D_{\rho\sigma}\left[2(\mu\nu|\rho\sigma) - (\mu\rho|\sigma\nu)\right] \tag{24}$$

At the onset of solving the Hartree-Fock equations, neither the orbitals nor the energies are known. However, the matrix equation (20) can be solved iteratively. A basis set is chosen and a guess is made at the orbital coefficient matrix, *C*. From this, the orbital coefficient density matrix *D* can be computed. The Fock matrix is then constructed via Eq. (24) and diagonalized. Since the energy matrix ε is diagonal, the diagonalization of the Fock matrix gives a new set of orbital coefficients, which can then be used to compute a new density matrix. The process continues until the density matrix ceases to change (or, in practical terms, until the difference between the current and previous iterations of the density matrix differ by less than a predetermined value).

The many-electron wavefunction is an antisymmetrized product of the one-electron Hartree-Fock orbitals. As shown in Eq. (14), the wavefunction must be antisymmetric; correct symmetry is obtained by constructing a normalized linear combination of Hartree products. This linear combination can be expressed conveniently in the form of a Slater determinant

$$\psi_{elec} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(1) & \phi_2(1) & \ldots & \phi_k(1) \\ \phi_1(2) & \phi_2(2) & \ldots & \phi_k(2) \\ \ldots & \ldots & & \ldots \\ \phi_1(N) & \phi_2(N) & \ldots & \phi_k(N) \end{vmatrix} \tag{25}$$

where N in the normalization constant is the total number of electrons, k is the number of spin-orbitals ϕ, and the number in parentheses is the number of the electron (i.e. electron i having coordinates $\vec{r}_i$).

Hartree-Fock is a relatively simple, computationally inexpensive *ab initio* method, scaling as $O(N^4)$ in the number of basis functions. Additionally, it has the advantage of being variational: the Hartree-Fock energy obtained from any given set of molecular orbitals is always an upper bound to the true energy that would be obtained if the full electronic Schrödinger equation were solved for the system. Since no set of orbitals can give an energy that is lower than the true energy, orbitals can be refined to minimize the energy. The minimum Hartree-Fock energy for a chemical system, as the basis set approaches completeness, is the Hartree-Fock limit.

Since it is a very basic approximation method, however, Hartree-Fock fails to describe many chemical systems accurately because it does not fully account for the correlated motion of electrons. An accurate depiction of electron correlation is lost when the mean field approximation is employed, rather than explicitly treating the interaction of each electron with every other electron. Specifically, while Hartree-Fock theory accounts for Fermi correlation (i.e. electron exchange) between electrons of like spin, it entirely neglects the Coulomb correlation between electrons with opposite spin. Higher levels of theory must be used to recover some of this electron correlation energy.

One of the most common methods utilized to go beyond the Hartree-Fock approximation is second-order Møller-Plesset perturbation theory[12] (MP2). In this method, electron correlation is treated as a small perturbation to a Hartree-Fock-like Hamiltonian:

$$\hat{H} = \hat{H}_0 + \lambda \hat{V} \tag{26}$$

Here, the MP2 Hamiltonian $\hat{H}$ is expressed as the sum of the unperturbed (zeroth-order) Hartree-Fock Hamiltonian $\hat{H}_0 = \sum_i \hat{F}(i) = \sum_i \left[\hat{h}(i) + \hat{v}^{HF}(i) \right]$ and a perturbation $\hat{V} = \sum_{i<j} r_{ij}^{-1} - \sum_i v^{HF}(i)$ representing the correlation potential. $\hat{V}$ is multiplied by a parameter, λ, that can vary from 0 (no perturbation) to 1 (full perturbation). The Schrödinger equation with the perturbed Hamiltonian is then

$$(\hat{H}_0 + \lambda\hat{V})\psi_n = E_n\psi_n \tag{27}$$

The energy and wavefunction can be written in the form of a power series in λ:

$$E_n = E_n^{(0)} + \lambda E_n^{(1)} + \frac{\lambda^2 E_n^{(2)}}{2} + ... + \frac{\lambda^k E_n^{(k)}}{k!} + ... \tag{28}$$

$$\psi_n = \psi_n^{(0)} + \lambda\psi_n^{(1)} + \frac{\lambda^2\psi_n^{(2)}}{2} + ... + \frac{\lambda^k\psi_n^{(k)}}{k!} + ... \tag{29}$$

where the superscripts in parentheses indicate the order of the derivative in λ. When λ=0, Eq. (28) and (29) reduce to the unperturbed values, leaving only the first term in each series (the "zeroth" derivative). When the values for E_n and ψ_n given by Eq. (28) and (29) are inserted into the Schrödinger equation (27), the resulting terms can be grouped into an infinite series of simultaneous equations based on like powers of λ. The zeroth-order equation (the collected terms that are multiplied by $\lambda^0 = 1$) is simply the unperturbed Schrödinger equation, Eq. (3), with ground state energy $E_0^{(0)}$ and wavefunction $\psi_0^{(0)}$. The first-order terms are grouped to obtain expressions for $E_n^{(1)}$ and $\psi_n^{(1)}$ in terms of the zeroth-order energy and wavefunction, then the second-order terms can be grouped to obtain expressions for $E_n^{(2)}$ and $\psi_n^{(2)}$ in terms of the first-order – and, in turn, zeroth-order – energies and wavefunctions.

The first-order energy correction is the expectation value of the perturbation operator, i.e.

$$E_0^{(1)} = \left\langle \psi_0^{(0)} \middle| \hat{V} \middle| \psi_0^{(0)} \right\rangle = \int \left(\psi_0^{(0)}(\vec{x})\right)^* \hat{V}\psi_0^{(0)}(\vec{x})d\vec{x} \tag{30}$$

The sum of the zeroth-order energy and the first-order energy correction gives the Hartree-Fock energy, so perturbation theory corrections to Hartree-Fock begin with the second-order terms. The second-order correction to the ground state energy is given by

$$E_0^{(2)} = \sum_{n>0} \frac{\left|\left\langle \psi_0^{(0)} \middle| V \middle| \psi_n^{(0)} \right\rangle\right|^2}{E_0^{(0)} - E_n^{(0)}} \tag{31}$$

The summation in Eq. (31) goes over all states but the ground state.

Spin component scaled second-order perturbation theory[13-15] (SCS-MP2) is a semi-empirical adaptation of MP2 that has shown greater accuracy in modeling some chemical systems for which MP2 is known to be in error. The method's origins lie in the fact that Hartree-Fock (equivalent to MP1) uses an unbalanced treatment of electron pairs that have

the same spin versus those with opposite spin. As previously discussed, Hartree-Fock electron correlation accounts for Fermi correlation (between electrons of like spin), but entirely neglects Coulomb correlation (between electrons of opposite spin). Building on this unbalanced description causes MP2 to give a biased treatment of short-range versus long-range correlation effects. In particular, MP2 often overestimates the correlation of distant electrons with the same spin. SCS-MP2 attempts to correct for this source of error by separating the second-order energy correction into components dealing with parallel spins and those dealing with anti-parallel spins, multiplying each by a different, empirically-determined scaling factor:

$$E_{SCS}^{(2)} = p_T\left(\sum e_{\alpha\alpha}^{(2)} + \sum e_{\beta\beta}^{(2)}\right) + p_S \sum e_{\alpha\beta}^{(2)} \tag{32}$$

where α and β are the possible spins of interacting electrons e and the scaling factors p_T and p_S have the values $^1/_3$ and $^6/_5$, respectively.

Resolution of the identity perturbation theory (RI-MP2),[16,17] another variant of MP2, employs approximations to reduce the most computationally demanding four-center two-electron repulsion integrals to combinations of less computationally expensive two- and three- center, one- and two-electron integrals.

Second-order perturbation theory is often the method of choice for *ab inito* quantum mechanical calculations; it accounts for electron correlation reasonably well in many chemical systems without incurring an unbearable computational cost. (MP2 scales as $O(N^5)$ in the number of basis functions.) However, because the exact Hamiltonian is not used, perturbation theory is not variational. Additionally, while it can be extended to higher orders of energy corrections (MPn, where n is an integer), it may not converge for larger n.

Coupled cluster theory[18-21] is a more computationally expensive, but generally even more accurate, alternative to MP2. The coupled cluster wavefunction is a linear combination of Hartree-Fock ground state and excited state Slater determinants; the excited state determinants account for the electron correlation. The coupled cluster wavefunction is expressed as

$$|\psi\rangle = e^{\hat{T}}|\Phi_0\rangle \tag{33}$$

where $|\Phi_0\rangle$ is the ground state Slater determinant of Hartree-Fock molecular orbitals and $\hat{T}$ is the cluster operator, which is the sum of the single excitation operator ($\hat{T}_1$), double excitation operator ($\hat{T}_2$), triple excitation operator ($\hat{T}_3$), and so on.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ... \tag{34}$$

$e^{\hat{T}}$ can be expanded in a Taylor series as

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + ... = 1 + \hat{T}_1 + \hat{T}_2 + \frac{\hat{T}_1^2}{2} + \hat{T}_1\hat{T}_2 + \frac{\hat{T}_2^2}{2} + ... \tag{35}$$

The excitation operators take the form

$$\hat{T}_1 = \sum_i \sum_a t_i^a \hat{a}_i \hat{a}_a^T \tag{36}$$

$$\hat{T}_2 = \frac{1}{4} \sum_{i,j} \sum_{a,b} t_{ij}^{ab} \hat{a}_i \hat{a}_j \hat{a}_a^T \hat{a}_b^T \tag{37}$$

etc., where $\hat{a}^T$ and $\hat{a}$ are the "creation" and "annihilation" operators, *i* and *j* are occupied orbitals, and *a* and *b* are virtual (unoccupied) orbitals. The effect of these operators is to "annihilate" one or more of the electrons in the occupied orbitals and "create" a corresponding number of electrons in the virtual orbitals. It is necessary to solve for the coefficients t_i^a and t_{ij}^{ab} during the process of finding $|\psi\rangle$. This is done by solving the set of equations

$$\langle \psi_0 | e^{-(\hat{T}_1 + \hat{T}_2 + ...)} \hat{H} e^{(\hat{T}_1 + \hat{T}_2 + ...)} | \psi_0 \rangle = E \tag{38}$$

$$\langle \psi_S | e^{-(\hat{T}_1 + \hat{T}_2 + ...)} \hat{H} e^{(\hat{T}_1 + \hat{T}_2 + ...)} | \psi_0 \rangle = 0 \tag{39}$$

$$\langle \psi_D | e^{-(\hat{T}_1 + \hat{T}_2 + ...)} \hat{H} e^{(\hat{T}_1 + \hat{T}_2 + ...)} | \psi_0 \rangle = 0 \tag{40}$$

and so on, where ψ_0 is the ground state wavefunction, ψ_S is the wavefunction of single excitations, ψ_D is the wavefunction of double excitations, etc. The cluster operator (and the number of equations) is truncated at the desired number of excitations. The "gold standard"[22] of computational chemistry is coupled cluster with singles, doubles, and perturbatively-calculated triple excitations – CCSD(T)[23]. CCSD(T) scales as $O(N^7)$ in the number of basis functions.

While not used in any of the following chapters, it is worth noting that all of the electron correlation energy could be recovered (the exact solution to the time-independent

Schrödinger equation, within the Born-Oppenheimer approximation, could be obtained) if full configuration interaction[24,25] (CI) were used with an infinite basis set. The CI wavefunction is, similarly to the coupled cluster wavefunction, a linear combination of excited state wavefunctions:

$$\psi = \sum_{i}^{occ}\sum_{a}^{vir} c_i^a \Phi_i^a + \sum_{i<j}^{occ}\sum_{a<b}^{vir} c_{ij}^{ab} \Phi_{ij}^{ab} + \sum_{i<j<k}^{occ}\sum_{a<b<c}^{vir} c_{ijk}^{abc} \Phi_{ijk}^{abc} + \ldots \quad (41)$$

CI can be truncated after any term. Full CI excites all the electrons of the system into the virtual orbitals. CI is incredibly computationally expensive, scaling exponentially with the number of basis functions, so it is only a viable method for the smallest chemical systems.

High-level *ab initio* calculations are often the standard on which other computational methods are judged, especially when experimental data is not available for a particular chemical system. Chapters 3 and 4 describe studies in which MP2, SCS-MP2, and CCSD(T) calculations have provided virtually the only data available for the systems of interest (non-equilibrium states of benzene, substituted benzene, and pyridine dimers). High-level *ab initio* calculations may also aid in interpreting ambiguous experimental (spectroscopic) data, or mediate between conflicting results found with lower-level calculations. Such is the case presented in Chapter 2, an investigation of the structure of small aluminum clusters. Finally, *ab initio* methods are the foundation of the field of computational chemistry. As such, the language, concepts, and theories associated with *ab initio* methods necessarily touch all work in this field in some fashion.

Density functional theory

Density functional theory (DFT) is a popular alternative to *ab inito* methods for electronic structure studies. Unlike the methods discussed previously, DFT does not attempt to solve the Schrödinger equation. Instead, it begins with the idea that the energy of a chemical system is a functional (a function of another function) of the electron density, ρ, which is in turn a function of the coordinates of the nuclei. Because the electron density depends on only three spatial coordinates, this would eliminate the many-body problem associated with solving the Schrödinger equation. The premise behind DFT is sound: in 1964, Hohenberg and Kohn[26] showed that various important ground state properties of a

many-electron system are uniquely determined by a functional of the electron density, and that the correct ground state electron density minimizes this functional. However, the exact mathematical form of this functional is not known.

Kohn and Sham[27] introduced a DFT framework that resembles the Hartree-Fock method, which retains the concept of electron orbitals. The Kohn-Sham equations are given by

$$\left(-\frac{1}{2}\nabla^2 + v_{eff}(\vec{r})\right)\phi_i(\vec{r}) = \varepsilon_i\phi_i(\vec{r}) \tag{42}$$

where v_{eff} is the Kohn-Sham potential, ϕ_i is a Kohn-Sham orbital, and ε_i is the orbital energy. As with the Hartree-Fock approach, the Kohn-Sham wavefunction is a single Slater determinant of the orbitals that are the lowest energy solutions to Eq. (42). An electron density $\rho = \sum_i |\phi_i|^2$ can be calculated from the set of Kohn-Sham orbitals, from which the energy can then be determined. The energy functional E is given by

$$E[\rho] = T_S[\rho] + \int d\vec{r}\cdot v_{ext}(\vec{r})\rho(\vec{r}) + V_H[\rho] + E_{XC}[\rho] \tag{43}$$

where T_S is the Kohn-Sham kinetic energy functional, v_{ext} gives the external potential acting on the system (e.g. the electron-nuclear interaction), V_H is the Coulomb functional, and E_{XC} is the exchange-correlation functional. Equations for the kinetic energy and Coulomb functionals are given by

$$T_S[\rho] = \sum_{i=1}^{N}\int d\vec{r}\,\phi_i^*(\vec{r})\left(-\frac{1}{2}\nabla^2\right)\phi_i(\vec{r}) \tag{44}$$

and

$$V_H = \frac{1}{2}\int d\vec{r}\int d\vec{r}\,'\frac{\rho(\vec{r})\rho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} \tag{45}$$

However, the formula for the exchange-correlation functional is not known, except in the case of a free electron gas.

An exchange-correlation functional is often produced through at least semi-empirical means. Many different functionals exist; most involve fitted parameters. Approaches to constructing functionals can be divided into four classes: the local density approximation (LDA), the local spin-density approximation (LSDA), the generalized gradient

approximation (GGA), and the meta-GGA. With LDA, the simplest type, the exchange-correlation energy is approximated as a function of the electron density at the coordinate where the functional is evaluated. LSDA extends LDA to include electron spin as a contributing factor. GGA takes into account the gradient of the electron density in addition to the value of the density itself. Meta-GGA methods, the most computationally costly, include a dependence on the second derivative of the density as well. Some functionals, called hybrid functionals, use all or a portion of the Hartree-Fock exchange with additional exchange and correlation energy included by one of the four methods listed above.

DFT can be as computationally inexpensive as Hartree-Fock while mirroring the accuracy of MP2, at least in systems similar to those with which the fitted parameters were determined. However, it can be worse than Hartree-Fock if an improper functional is chosen or if a good functional does not exist for the chemical system of interest. Different DFT functionals may provide different results for calculations on the same system, as can be seen in the aluminum cluster study in Chapter 2.

Modeling intermolecular interactions

Intermolecular interactions are significantly weaker than the chemical bond, and are correspondingly difficult to capture. While chemical bonds are on the order of one hundred kilocalories per mole, minima on the intermolecular potential energy surface may reach depths from mere tenths or hundredths of a kilocalorie per mole to several kilocalories per mole. Due to this energy discrepancy, most of the common methods used for electronic structure calculations do not perform well at describing intermolecular interactions. Standard electronic structure methods are used with the supermolecular approach, in which interaction energy is computed by subtracting the individual monomers' energies from the energy of the combined system:

$$E_{\text{int}} = E_{AB} - (E_A + E_B) \tag{46}$$

This approach is problematic due to basis set superposition error[28-30] (BSSE). An infinite basis set can clearly never be achieved, but using a finite basis set means that a greater total number of basis functions is available to each monomer when a calculation is performed on a dimer system than when a calculation is performed on each monomer separately. A monomer

"borrows" from the basis set of another monomer if it is close enough to interact. This results in an artificial lowering of the energy of the dimer system relative to that of the individual monomers. While various methods exist that attempt to calculate, and hence eliminate, the BSSE, there is no consensus on how best to do so. Additionally, without using a very high-level standard electronic structure method, errors in the total energies may exceed the magnitude of the interaction energy itself. The supermolecular approach relies on cancellation of (large) errors, which may not always occur as desired.

In addition to the problems associated with obtaining an accurate intermolecular interaction energy with the supermolecular approach, that approach does not offer desired physical insight into the nature of the interaction. Intermolecular interactions are typically divided into four fundamental contributions: Coulomb (electrostatic), induction (polarization), dispersion, and exchange-repulsion. The Coulomb term results from interactions between the permanent electric multipole moments on different monomers. Induction arises from interactions between permanent electric multipole moments on one monomer and induced multipole moments on the other. The dispersion energy results from interactions between instantaneous electric multipoles. Exchange refers to the interchanging of electrons between monomers due to tunneling. The strongest of these interactions is often the Coulomb term – for example, in hydrogen bonded systems. However, there are also chemical systems stabilized primarily by the dispersion interaction (a focus of most of this dissertation). A fifth contribution to intermolecular interactions, important in systems with charged or highly polar species, is charge transfer, which arises from the interaction between occupied orbitals on one molecule and virtual orbitals on another.

Given the difference in magnitude between the energies of monomers and the energies of intermonomer interactions, it is intuitive to treat the relatively small intermonomer interaction with perturbation theory. The perturbative Hamiltonian $\hat{H}$ can be constructed as in Eq. (26)

$$\hat{H} = \hat{H}_0 + \lambda\hat{V}$$

and the unperturbed Hamiltonian $\hat{H}_0$ can be partitioned as

$$\hat{H}_0 = \hat{H}_A + \hat{H}_B \tag{47}$$

where $\hat{H}_A$ and $\hat{H}_B$ are the unperturbed electronic Hamiltonians for monomers A and B. The perturbation operator $\hat{V}$ then corresponds to all Coulombic interactions between the electrons and nuclei of monomer A on one side and the electrons and nuclei of monomer B on the other. Hence, in this context, it is often called the intermolecular interaction operator. When the perturbation is turned off ($\lambda = 0$), the ground state unperturbed energy is a sum of monomer energies

$$E_0^{(0)} = E_{AB}^{(0)} = E_A + E_B \tag{48}$$

and the corresponding wavefunction is the product of the individual monomer wavefunctions

$$\psi_0^{(0)} = \psi_{AB}^{(0)} = \psi_A \psi_B \tag{49}$$

When the perturbation is turned on ($\lambda = 1$), the intermolecular interaction terms in $\hat{V}$ are completely included, so the corresponding wavefunction ψ_{AB} and energy E_{AB} become the exact wavefunction and energy of the dimer. Thus, ψ_{AB} and E_{AB} are more properly expressed as functions of the perturbation parameter λ.

The wavefunction ψ_A is generally given as a function of only the electrons on monomer A, and the Hamiltonian $\hat{H}_A$ acts solely on these electrons; likewise for ψ_B and $\hat{H}_B$ with the electrons on monomer B. The problem with this approach is that, except at long distances, electrons can tunnel between monomers. The correct unperturbed wavefunction must account for this electron exchange. The antisymmetrization operator $\hat{A}$ can be applied to the unperturbed wavefunction to give correct parity under electron exchange. This operator is given by

$$\hat{A} = \frac{1}{\sqrt{N!}} \sum (-1)^{\pi} \hat{P} \tag{50}$$

where N denotes the number of electrons, the sum goes over all possible instances of $\hat{P}$, π is the number of electron exchanges occurring for each instance of $\hat{P}$, and $\hat{P}$ itself is the transposition operator. $\hat{P}$ can be expanded as a sum of permutation operators that exchange two, three, four, etc. electrons. The two-electron permutation operator is given by:

$$\hat{P}_{ij}\psi(1,2,...,i,...,j,...,N) = \psi(1,2,...,j,...,i,...,N) \tag{51}$$

$\hat{A}\psi_0^{(0)}$ gives the symmetry-corrected unperturbed zeroth-order wavefunction; however, this function is no longer an eigenvalue of the unperturbed Hamiltonian. The

perturbation procedure must be modified so that $\hat{A}\psi_0^{(0)}$ can be used with the desired, physically intuitive construction of the Hamiltonian, $\hat{H}_0 = \hat{H}_A + \hat{H}_B$. These various modifications are known as symmetry adaptation, and the resulting procedure as symmetry adapted perturbation theory[31-33] (SAPT). So far, the most successful symmetry adaptation has been in "weak symmetry forcing," in which the antisymmetrizer is used only to correct the energy expressions; otherwise, perturbation theory is carried out on the (unsymmetrized) approximate wavefunction $\psi_0^{(0)}$. The corrected (symmetrized) interaction energy expression becomes

$$E_{\text{int}}(\lambda) = \frac{\left\langle \psi_0^{(0)} \middle| \lambda \hat{V} \middle| \hat{A}\psi_{AB}(\lambda) \right\rangle}{\left\langle \psi_0^{(0)} \middle| \hat{A}\psi_{AB}(\lambda) \right\rangle} \tag{52}$$

E_{int} can be expanded in powers of λ as

$$E_{\text{int}}^{(n)} = \frac{\left\langle \psi_0^{(0)} \middle| \hat{V} \middle| \hat{A}\psi_{AB}^{(n-1)} \right\rangle}{\left\langle \psi_0^{(0)} \middle| \hat{A}\psi_0^{(0)} \right\rangle} - \sum_{k=1}^{n-1} E_{\text{int}}^{(k)} \left\langle \psi_0^{(0)} \middle| \hat{A}\psi_{AB}^{(n-k)} \right\rangle \tag{53}$$

The n-th order interaction energy is the sum of the n-th order reference energy term and a short-range correction that depends on the portion of the antisymmetrizer that exchanges electrons between monomers:

$$E_{\text{int}}^{(n)} = E_{ref}^{(n)} + E_{exch}^{(n)} \tag{54}$$

The energy terms on the right side of Eq. (54) can be decomposed into the familiar components of the interaction energy. The Coulomb interaction energy term is identical to $E_{ref}^{(1)}$. $E_{ref}^{(2)}$ is equivalent to the sum of the induction and dispersion interaction terms. If carried out to third order, $E_{ref}^{(3)}$ can be decomposed into an induction term, a dispersion term, and an induction-dispersion term resulting from the coupling of the induction and dispersion interactions; the third-order induction and dispersion terms may be added to the respective second-order terms as corrections. The first-order exchange energy is simply $E_{exch}^{(1)}$. Similarly to $E_{ref}^{(2)}$, the second-order expression $E_{exch}^{(2)}$ can be separated into exchange-induction and exchange-dispersion contributions.

It is convenient to begin with reference wavefunctions for each monomer computed via Hartree-Fock, which will then need to be corrected for electron correlation within

monomers while being used to compute intermolecular interactions. In this case, two perturbation operators are used simultaneously: $W = W_A + W_B$ accounts for intramonomer correlation (within monomers A and B), while V accounts for intermonomer interactions (as shown previously). The double perturbation expression, beginning with a Hartree-Fock wavefunction $\psi_{AB}(0,0) = \psi_0^{HF}$, is

$$(\hat{F} + \zeta\hat{W} + \lambda\hat{V})\psi_{AB}(\zeta,\lambda) = E_{AB}(\zeta,\lambda)\psi_{AB}(\zeta,\lambda) \tag{55}$$

where $\hat{F} = \hat{F}_A + \hat{F}_B$ is the sum of the Fock operators for monomers A and B, and ζ and λ are equal to 1 when the perturbations are "on." As with the single perturbation expression shown previously, the energy and wavefunction of Eq. (55) can be expanded in powers of ζ and λ; grouping by like powers produces expressions for each order of the perturbation theory. For the energy, this yields doubly-ordered terms resembling $E^{(nl)}$, where *n* and *l* denote the order of the perturbation in V and W, respectively. A list of the significant energy terms is given below[33,34].

$E_{elst}^{(10)}$	Coulomb (electrostatic) energy
$E_{elsr}^{(1l)}, l = 2,3,4$	intramonomer correlation corrections to Coulomb energy
$E_{exch}^{(10)}$	exchange-repulsion
$E_{exch}^{(1l)}, l = 1,2$	intramonomer correlation corrections to exchange-repulsion
$E_{ind}^{(20)}$	induction energy
$E_{ind}^{(22)}$	intramonomer correlation corrections to the induction energy
$E_{exch-ind}^{(20)}$	exchange-induction energy
$E_{disp}^{(20)}$	dispersion energy
$E_{disp}^{(2l)}, l = 1,2$	intramonomer correlation corrections to the dispersion energy

While SAPT is often the method of choice for calculating intermolecular interactions, as an *ab initio* perturbation method, it is a time-consuming calculation. The effective fragment potential[35-37] (EFP) method is an *ab initio*-based method of calculating intermolecular interactions that is coded in the GAMESS[38] (General Atomic and Molecular Electronic Structure System) quantum chemistry software package. The EFP method was

originally developed as a discrete solvation method. Discrete solvation methods model all solvent molecules explicitly, rather than treating the entire solvent with a single model potential. Either the solute and solvent can both be treated with EFP, or the EFP method can be made to interface with an *ab initio* method. The bulk solvent can be modeled computationally inexpensively with EFP while the solute is treated with a higher-level method; interactions between the solute and the EFP solvent molecules are treated as perturbations to the solute Hamiltonian. EFP1 exclusively models solvation with water, having fitted parameters to model the exchange-repulsion. EFP2, the general EFP method, can be used to model any type of monomer and has no fitted parameters. EFP2 will be the focus of the remainder of this section.

An EFP2 calculation is run in two parts. First, the necessary parameters are generated in a calculation called MAKEFP in GAMESS. This step needs to be performed only once for each type of monomer. (It can be bypassed if the parameters are already available in an EFP library.) Once the parameters that describe the effective fragment have been calculated, they can be used again and again in a RUNEFP calculation.

The EFP method computes Coulomb, induction (polarization), dispersion, and exchange-repulsion terms. The Coulomb potential at a point $\vec{r}_2$ can be expressed in terms of multipoles centered around a point $\vec{r}_1$ via the expression

$$V_{\vec{r}_1}^{Coul}(\vec{r}_2) = \frac{q^{\vec{r}_1}}{\vec{r}_{12}} + \sum_{\alpha}^{x,y,z} \mu_{\alpha}^{\vec{r}_1} F_{\alpha}(\vec{r}_{12}) + \frac{1}{3}\sum_{\alpha,\beta}^{x,y,z} \Theta_{\alpha\beta}^{\vec{r}_1} F_{\alpha\beta}(\vec{r}_{12}) - \frac{1}{15}\sum_{\alpha,\beta,\gamma}^{x,y,z} \Omega_{\alpha\beta\gamma}^{\vec{r}_1} F_{\alpha\beta\gamma}(\vec{r}_{12}) + \ldots \qquad (56)$$

where q, μ, Θ, and Ω are the charge, dipole, quadrupole, and octopole, respectively, and F_{α}, $F_{\alpha\beta}$, and $F_{\alpha\beta\gamma}$ are the electric field, field gradient, and field second derivatives. The expression in Eq. (56) converges to the exact Coulomb potential when an infinite number of terms are included. Because the expression converges slowly, distributed multipole analysis[34] (DMA), in which many points are used as centers of the multipole expansion, is preferable. Thus, Eq. (56) is expanded in a series in $\vec{r}_{12}^{-1}$ to obtain the DMA Coulomb energy expression

$$E_{ij}^{Coul} = q^j\left[Tq^i - T_{\alpha}\mu_{\alpha}^i + \frac{1}{3}T_{\alpha\beta}\Theta_{\alpha\beta}^i - \frac{1}{15}T_{\alpha\beta\gamma}\Omega_{\alpha\beta\gamma}^i\right]$$
$$+\mu_{\alpha}^j\left[T_{\alpha}q^i - T_{\alpha\beta}\mu_{\beta}^i + \frac{1}{3}T_{\alpha\beta\gamma}\Theta_{\beta\gamma}^i - \frac{1}{15}T_{\alpha\beta\gamma\delta}\Omega_{\beta\gamma\delta}^i\right] \qquad (57)$$

$$+\Theta^B_{\alpha\beta}\left[T_{\alpha\beta}q^A - T_{\alpha\beta\gamma}\mu^A_\gamma + \frac{1}{3}T_{\alpha\beta\gamma\delta}\Theta^A_{\gamma\delta} - ...\right] + ...$$

for the interaction between two EFP2 monomers. In Eq. (56), the sums over the directional components (given by Greek letters) are implicit. The Coulomb interaction between each pair of expansion points (*i* on monomer *A* and *j* on monomer *B*) is computed as the interaction between the pairs of multipoles, up to octopoles. (The EFP2 Coulomb calculation is truncated at octopole interactions.) The quantities T, T_α, $T_{\alpha\beta}$, etc. are electrostatic tensors of the zeroth-, first-, second-, etc. order. The first three tensors are given by

$$T = \frac{1}{R} \tag{58}$$

$$T_\alpha = \nabla_\alpha \frac{1}{R} = -\frac{R_\alpha}{R^3} \tag{59}$$

$$T_{\alpha\beta} = \frac{3R_\alpha R_\beta - R^2\delta_{\alpha\beta}}{R^5} \tag{60}$$

The terms in Eq. (57) can be grouped and expressed as

$$E_{ij}^{Coul} = E_{ij}^{ch-ch} + E_{ij}^{ch-dip} + E_{ij}^{ch-quad} + E_{ij}^{ch-oct} + E_{ij}^{dip-dip} + E_{ij}^{dip-quad} + ... + E_{ij}^{oct-oct} \tag{61}$$

In EFP2, atom centers and bond midpoints (points with high electron density) are used as the multipole expansion points. The charge, dipole, quadrupole, and octopole terms for each EFP2 monomer type need only be calculated once, in a MAKEFP calculation.

Interactions between each nucleus *I* and the multipoles centered at *j* are also computed and summed:

$$E_{Ij}^{Coul} = E_{Ij}^{nuc-ch} + E_{Ij}^{nuc-dip} + E_{Ij}^{nuc-quad} + E_{Ij}^{nuc-oct} \tag{62}$$

The Coulomb interaction between nucleus *I* and nucleus *J* has the classical form

$$E_{IJ}^{Coul} = \frac{Z_I Z_J}{R_{IJ}} \tag{63}$$

where Z_I is the nuclear charge. The total EFP2 Coulomb energy term between monomers *A* and *B* includes all of these interactions:

$$E^{Coul} = \frac{1}{2}\sum_A \sum_{B\neq A}\left(\sum_{i\in A}\sum_{j\in B} E_{ij}^{Coul} + \sum_{I\in A}\sum_{J\in B} E_{IJ}^{Coul} + \sum_{I\in A}\sum_{j\in B} E_{Ij}^{Coul}\right) \tag{64}$$

The Coulomb interaction between an EFP2 monomer and an *ab initio* monomer is given as a perturbation to the *ab initio* Hamiltonian, as in Eq. (26). The perturbation operator $\hat{V}$ consists of contributions from all multipole expansion points and from all nuclei on all of the EFP2 monomers.

$$\hat{V} = \sum_m \sum_n \sum_A \left(\sum_{i \in A} D_{pq} \langle m | V_i^{Coul} | n \rangle + \sum_{I \in A} \langle m | \frac{Z_I}{R} | n \rangle \right) \tag{65}$$

Here, A is an EFP2 monomer, I is a nucleus, m and n are atomic orbitals on the *ab initio* monomer, Z_I is the nuclear charge, D_{pq} is an element of the density matrix, and V_i^{Coul} gives the Coulomb potential at an expansion point. The integral with V_i^{Coul} looks like

$$\begin{aligned} \langle m | V_i^{Coul} | n \rangle = \langle m | \frac{q^i}{R} | n \rangle + \langle m | \frac{\sum_\alpha^{x,y,z} \mu_\alpha^i \alpha}{R^3} | n \rangle + \langle m | \frac{\sum_{\alpha\beta}^{x,y,z} \Theta_{\alpha\beta}^i \left(3\alpha\beta - R^2 \delta_{\alpha\beta} \right)}{3R^5} | n \rangle \\ + \langle m | \frac{\sum_{\alpha\beta\gamma}^{x,y,z} \Omega_{\alpha\beta\gamma}^i \left(5\alpha\beta\gamma - R^2 (\alpha\delta_{\beta\gamma} + \beta\delta_{\alpha\gamma} + \gamma\delta_{\alpha\beta}) \right)}{5R^7} | n \rangle \end{aligned} \tag{66}$$

where R, α, β, and γ are the distance and the components of the distance between the electron and the expansion point i and $\delta_{\alpha\beta}$ is the Kronecker delta.

The EFP2 Coulomb interaction model described above performs well at large intermonomer separations, but performs more poorly when the electron densities of the monomers overlap. This overlap is known as charge penetration. Charge penetration in EFP2 is modeled by multiplying the charge-charge interaction term by a screening function.[39]

$$E_{ij}^{ch-ch} \rightarrow \begin{cases} \left(1 - \left(1 + \frac{\alpha_i R_{ij}}{2} \right) e^{-\alpha_i R_{ij}} \right) \frac{q_i q_j}{r_{ij}} & (if \quad \alpha_i = \alpha_j) \\ \left(1 - \frac{\alpha_i^2}{\alpha_i^2 - \alpha_j^2} e^{-\alpha_i R_{ij}} - \frac{\alpha_j^2}{\alpha_j^2 - \alpha_i^2} e^{-\alpha_i R_{ij}} \right) \frac{q_i q_j}{r_{ij}} & (if \quad \alpha_i \neq \alpha_j) \end{cases} \tag{67}$$

Screening may be applied to higher multipole interaction terms as well. The parameters α in the screening function are calculated to minimize the differences between the screened EFP2 electrostatic potential and the *ab initio* electrostatic potential for the same monomer, which is constructed during a MAKEFP calculation.

The quantum mechanical formula for induction (polarization), which describes a molecule's charge redistribution in the presence of an external electric field, is part of a

second-order term appearing in long-range perturbation theory. The equation for the induction energy of monomer *B* resulting from monomer *A* is

$$E^{ind;B} = -\sum_{n\neq 0} \frac{\langle 00|V^{Coul}|0n\rangle - \langle 0n|V^{Coul}|00\rangle}{E_n^B - E_0^B} \tag{68}$$

This is a sum-over-states formula in which *n* represents an excited electronic state and 0 represents the ground state, $|mn\rangle$ refers to a state *m* on monomer *A* and a state *n* on monomer *B*, and E_n is the energy of state *n*. Using the multipole expansion of Eq. (57) for the perturbation operator in Eq. (67) and simplifying,

$$E^{ind;B} = -\sum_{i\in A}\sum_{j\in B}\left(q^i T_a - \mu_b^i T_{ab} + ...\right)\sum_{n\neq 0} \frac{\langle 0|\mu_a^j|n\rangle\langle n|\mu_{a'}^j|0\rangle}{E_n^B - E_o^B}\left(q^i T_{a'} - \mu_{b'}^i T_{a'b'} + ...\right) \tag{69}$$

The expression in Eq. (69) omits the quadrupole and higher terms because EFP2 uses only the lower-order terms for computational efficiency. The distributed approach allows reasonable accuracy to be retained even when the higher-order multipoles are neglected, because of the use of localized molecular orbitals (LMOs). Labeling the summation in the center of Eq. (69) as a polarizability $\alpha_{aa'}$ and noting that $\left(q^i T_a - \mu_b^i T_{ab} + ...\right)$ corresponds to minus the electric field at *j* due to *i*, the final induction energy expression is

$$E^{ind;B} = -\sum_{i\in A}\sum_{j\in B} F_a^i(j) F_{a'}^i(j) \alpha_{aa'}^j \tag{70}$$

Induced dipoles are calculated at each polarizability point (in EFP2, these correspond to bonds and lone pairs) using the expression

$$\mu_i = \alpha_i F_i \tag{71}$$

where the terms μ_i, α_i, and F_i give the induced dipole, the polarizability tensor, and the external field at point *i*. (While the polarizability tensor for the entire molecule is symmetric, the distributed polarizability tensors are asymmetric, so all nine components are needed.) The induced dipoles are used to calculate the induction energy

$$E^{ind} = \sum_i \mu_i F_i \tag{72}$$

The field itself is dependent on the permanent multipoles, induced dipoles, and nuclei of other EFP2 monomers; in the case of a mixed *ab initio*-EFP2 system, it depends on the electron density and nuclei of the *ab initio* molecule as well. The induced dipoles are affected

by the field created by other induced dipoles or by the *ab initio* electron density, which is in turn affected by the field created by the induced dipoles through a one-electron contribution to the Hamiltonian

$$H^{ind} = \sum_{A} \sum_{i \in A} \sum_{a}^{x,y,z} \frac{(\mu_{i,a} + \overline{\mu}_{i,a})a}{R^3} \tag{73}$$

R gives the distance between the electron and the polarizability point i, a gives the distance components, $\mu_{i,a}$ denotes the ath distance component of the induced dipole at location i, and $\overline{\mu}_{i,a}$ is the conjugated induced dipole, which is computed with

$$\overline{\mu}_i = \alpha_i^T F_i \tag{74}$$

where α_i^T is the transpose of the polarizability tensor. Because of these intertwined dependencies, the induced dipoles must be computed self-consistently. If the entire system is modeled with EFP2, only the induced dipoles and external fields must be made self-consistent; in a mixed *ab initio*-EFP2 system, the *ab initio* wavefunction must be consistent with the induced dipoles and fields as well.

The polarization parameters determined in a MAKEFP calculation are polarizability tensors and LMOs. LMOs are obtained from the *ab initio* electron density using the Boys[40] localization method. Polarizability tensors are obtained by a finite difference method based on the relation

$$\alpha_{ab} = \frac{\partial \mu_a}{\partial F_b} \tag{75}$$

First, a dipole moment is computed for each LMO in the absence of an electric field. Then, in three additional calculations, the dipole moments are computed in the presence of a small electric field acting along each coordinate axis. Finally, the polarizability tensor is computed as

$$\alpha_{ab} = \frac{\mu_a(F_b) - \mu_a(0)}{F_b} \tag{76}$$

Along with the induction term, the quantum mechanical expression for the dispersion energy appears as part of a second-order term in long-range perturbation theory. The sum-over-states expression for the dispersion energy between interacting molecules A and B is given by

$$E^{disp} = -\sum_{n\neq 0}\sum_{m\neq 0}\frac{\langle 00|V^{Coul}|mn\rangle\langle mn|V^{Coul}|00\rangle}{E_m^A + E_n^B - E_0^A - E_0^B} \tag{77}$$

where 0 is the ground state and m and n are excited states with corresponding energies E_0, E_m, and E_n. When the multipole expansion is used, Eq. (76) can be expressed in the form

$$E^{disp} = \frac{C_6}{R^6} + \frac{C_8}{R^8} + \frac{C_{10}}{R^{10}} + \ldots \tag{78}$$

The coefficients C_n arise from the induced dipole-induced dipole interactions (C_6), induced dipole-induced quadrupole interactions (C_8), induced quadrupole-induced quadrupole and induced dipole-induced octopole interactions (C_{10}), and so on. The dispersion term is computed using a distributed model with expansion points at the LMO centroids.

The induced dipole-induced dipole interaction energy between dispersion points (LMOs) i and j can be expressed as[41,42]

$$E_{ij} = \sum_{abcd}^{x,y,z} T_{ab}T_{cd}\int_0^\infty \alpha_{ac}^i(i\nu)\alpha_{bd}^j(i\nu)d\nu \tag{79}$$

where a, b, c, and d give directional components, α_{ac}^i and α_{bd}^j are the dynamic polarizability tensors associated with each LMO, and the integral goes over the imaginary frequency range ($i\nu$). This integral over frequency produces the C_6 coefficient in Eq. (78). T_{ab} and T_{cd} are quadrupole tensors, given by Eq. (60). Details about the calculation of Eq. (79) can be found in Chapter 6.

In the expression for the total EFP2 dispersion energy, only the first term in Eq. (78) is calculated explicitly. It has been shown that the remainder of the series can be fairly well approximated as 1/3 of the $\frac{C_6}{R^6}$ value. Hence, the final expression becomes

$$E^{disp} = \frac{4}{3}\sum_A\sum_B\sum_{i\in A}\sum_{j\in B}\frac{C_6^{ij}}{R^6} \tag{80}$$

Damping of the EFP2 dispersion coefficients is performed to account for charge penetration. In the original formulation of EFP2, a Tang-Toennies[43,44] damping function was used. However, this function may overdamp the dispersion energy; a new damping function, dependent on the electronic density overlap matrix elements S^{ij} between LMOs i and j, is now preferred. This damping function is given by[45]

$$C_6^{ij} \rightarrow \left(1 - \left|S^{ij}\right|^2 \left(1 - 2\ln\left|S^{ij}\right| + 2\ln^2\left|S^{ij}\right|\right)\right) C_6^{ij} \tag{81}$$

The dispersion interaction between an EFP2 monomer and an *ab initio* molecule has not been implemented. Progress toward this goal is described in Chapter 6.

The dispersion parameters found in a MAKEFP calculation include the locations of the LMO centroids and values of the dynamic polarizability tensors for each LMO. The dynamic polarizability tensors are computed using the dynamic analog of the coupled perturbed Hartree-Fock equations[46].

The exchange-repulsion term[47,48] is the most computationally expensive EFP2 energy term discussed so far. As noted in the discussion of symmetry adapted perturbation theory (SAPT), exchange-repulsion results from the short-range overlap of individual molecular wavefunctions, where tunneling permits the interchange of electrons between molecules. As such, it is an entirely quantum mechanical effect. The full expression for the exchange-repulsion energy found with perturbation theory is given by

$$E^{exch} = \frac{\left\langle \psi_A \psi_B \left| \hat{A} \hat{H}_{AB} \right| \psi_A \psi_B \right\rangle}{\left\langle \psi_A \psi_B \left| \hat{A} \psi_A \psi_B \right\rangle} - \left\langle \psi_A \psi_A \left| \hat{V} \right| \psi_B \psi_B \right\rangle - E_A - E_B \tag{82}$$

where A and B denote the two (separated) molecules described by wavefunctions ψ_A and ψ_B with energies E_A and E_B, $\hat{H}_{AB} = \hat{H}_A + \hat{H}_B + \hat{V}$ is the unperturbed Hamiltonian, and $\hat{V}$ is the perturbation operator, which captures the total Coulomb interaction between the molecules. The antisymmetrization operator $\hat{A}$ is given in Eq. (50). Applying an infinite basis set approximation and a spherical gaussian overlap approximation[49], an expression for the exchange-repulsion in terms of LMOs is obtained:

$$E_{ij}^{exch} = -4\sqrt{\frac{-2\ln\left|S_{ij}\right|}{\pi}} \frac{S_{ij}^2}{R_{ij}} - 2S_{ij}\left(\sum_{k \in A} F_{ik}^A S_{kj} + \sum_{l \in B} F_{jl}^B S_{il} - 2T_{ij}\right)$$
$$+2S_{ij}^2\left(\sum_{J \in B} -Z_J R_{iJ}^{-1} + 2\sum_{l \in B} R_{il}^{-1} + \sum_{I \in A} -Z_I R_{Ij}^{-1} + 2\sum_{k \in A} R_{kj}^{-1} - R_{ij}^{-1}\right) \tag{83}$$

Here, A and B are the EFP2 monomers, i, j, k, and l are LMOs, I and J are nuclei, S is the the matrix of intermolecular overlap integrals, T is the matrix of kinetic energy integrals, and F is the Fock matrix.

While the EFP2 exchange-repulsion term is more expensive to compute than the EFP2 Coulomb, induction, or dispersion terms, it describes a short-range effect only. A proximity-based screening can determine for which pairs of EFP2 monomers it is necessary to compute this term. By eliminating the need to calculate the exchange-repulsion for every pair of EFP2 monomers, finding the total exchange-repulsion in a system of many monomers may be less computationally expensive than calculating the long-range terms, such as the total Coulomb interaction.

A final term, the charge-transfer interaction, exists in the EFP method. This interaction involves the promotion of electrons from the occupied orbitals on one molecule to the virtual orbitals on another. This term is very time-consuming to compute and only makes a significant contribution to the total interaction energy in systems of charged or highly polar molecules, such as water. For this reason, charge-transfer was not used in any of the work presented here. An overview of charge-transfer in EFP can be found in ref. 50.

The EFP2 method is the basis of most of this work. Chapters 3-5 are studies in which the EFP2 method is applied to dimer complexes that are normally difficult to model without resorting to computationally expensive, high-level calculations. Whenever possible in these chapters, EFP2 energy components are compared with SAPT values, because SAPT is a well-regarded method for determining the various contributions to the total interaction energy. Chapter 6 presents efforts toward extending the utility of the EFP2 method by implementing a dispersion energy term in mixed *ab initio*-EFP2 systems.

Exploring the potential energy surface

The potential energy for a chemical system is a function of a huge number of variables – 3N, where N is the number of particles of interest. These variables give the coordinates of each particle. For all but the simplest chemical systems, the number of degrees of freedom is too large to make an exhaustive enumeration feasible. Therefore, methods have been developed to sample the potential energy surface (PES) in search of interesting features. The most commonly sought point on the PES is the global minimum, but local minima and saddle points (especially first-order saddle points, which are transition states) are of interest as well. For example, these states may be intermediates in a chemical reaction.

This section focuses on Monte Carlo[51]/simulated annealing[52,53] (MC/SA), which repeatedly and randomly samples the PES. Random sampling methods are in contrast to optimization methods, which commonly use "greedy" algorithms. Greedy algorithms set a course along the PES determined by making the locally optimal choice at each step. Given a starting location on a PES, a greedy algorithm would, for example, compute the energy and gradient at that point and follow the trajectory of steepest descent to a lower energy. After following this trajectory a set distance, it computes the gradient at the new point and again travel along the steepest descent. This process continues until a minimum is reached. While this method does produce a sampling of the PES, it is possible for a greedy algorithm to become stuck in a local minimum that is higher in energy than the global minimum. Additionally, if a broader sampling of the PES is desired, the steeepest descent trajectory may be limiting.

MC/SA uses pseudo-random number generation (the "Monte Carlo" part of the name, eponymous with the famous casino) to choose the next move along the PES. Moving between states (locations on the PES) is also a function of a variable T, the "temperature." T begins at a (high) specified value and decreases in set increments with each iteration until reaching its minimum, another specified value. When T is high, the move is more likely to occur. The likelihood of moving progressively decreases as T approaches its minimum, at which point the program ends. Allowing greater movement in the beginning permits a broad sampling of the PES and, ideally, prevents the algorithm from converging on a final state that is a local, rather than the global, minimum.

The use of decreasing "temperature" comes from annealing in metallurgy, where defects in metals may be reduced by heating and then cooling them. Directly analogously with PES sampling in MC/SA programs designed for chemistry, the annealing of metals works by allowing atoms to move from their initial positions at sites of defects (local minima) to a lower energy state (ideally, the defect-free global minimum). Simulated annealing algorithms are not, however, limited to chemistry-related applications; they can be implemented to obtain good approximate solutions to many problems that have a large set of possible outcomes.

The disadvantage of MC/SA is that, due to its random sampling, many repetitions of the program may be needed to find the global minimum. However, the speed and simplicity of the MC/SA approach often make it feasible to conduct a large number of trials, especially when MC/SA is used with a fast method like EFP2 to compute the energy of each sampled geometry. Since the number of possible configurations decreases with system size, it is more likely that MC/SA will converge on the global minimum when small systems are examined. EFP2-MC/SA investigations of dimer potential energy surfaces appear in Chapters 3 and 4.

References

1) Schrödinger, E. *Ann. Physik.* **1926**, *79*, 361.
2) Schrödinger, E. *Ann. Physik.* **1926**, *79*, 489.
3) Schrödinger, E. *Ann. Physik.* **1926**, *79*, 734.
4) Schrödinger, E. *Ann. Physik.* **1926**, *80*, 437.
5) Schrödinger, E. *Ann. Physik.* **1926**, *81*, 109.
6) Born, M.; Oppenheimer, R. *Ann. Physik.* **1927**, *84*, 457.
7) Hartree, D. R. *Proc. Cambridge Phil. Soc.*, **1928**, *24*, 89.
8) Hartree, D. R. *Proc. Cambridge Phil. Soc.*, **1928**, *24*, 111.
9) Hartree, D. R. *Proc. Cambridge Phil. Soc.*, **1928**, *24*, 426.
10) Fock, V. *Physik* **1930**, *61*, 126.
11) Roothan, C. C. J. *Rev. Mod. Phys.*, **1960**, *32*, 179.
12) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
13) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
14) Grimme S., *J. Comput. Chem.* **2003**, *24*, 1529.
15) Gerenkamp, M.; Grimme, S. *Chem. Phys. Lett.* **2004**, *392*, 229.
16) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359.
17) Bernholdt, D. E.; Harrison, R. J. *Chem. Phys. Lett.* **1996**, *250*, 477.
18) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.
19) Paldus, J. *NATO ASI Series, Series B: Physics* **1992**, *293*, 99.
20) Bartlett, R. J. *Advanced Series in Physical Chemistry* **1995**, *2*, 1047.
21) Lee, T. J.; Scuseria, G. E. *Understanding Chemical Reactivity* **1995**, *13*, 47.
22) Gordon, M. S. Personal correspondence.
23) Piecuch, P.; Kucharski, S. A.; Kowalski, K.; Musial, M. *Comput. Phys. Commun.* **2000**, *149*, 71.
24) Shavitt, I. *The Method of Configuration Interaction*. Methods of Electronic Structure Theory; Schaefer, H. F., Ed.; Plenum Press: New York, **1977**, p. 189-275.
25) Sherrill, C.D.; Shaefer III, H. F. *Adv. Quant. Chem.*, **1999**, *34*, 143.
26) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.
27) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.
28) Jansen, H. B.; Ros, P. *Chem. Phys. Lett.* **1969**, *3*, 140.
29) Liu, B.; McLean, A. D. *J. Chem. Phys.* **1973**, *59*, 4557.
30) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

31) Jeziorski, B.; Kolos, W. In Molecular Interactions; Ratajczak, H; Orville-Thomas, W. J., Eds.; Wiley: New York, **1982**; Vol. 3, p. 1.
32) Rybak, S.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **1991**, *95*, 6576.
33) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.
34) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: New York/Oxford, **1996**.
35) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.
36) Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, *105*, 11081.
37) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V. *J. Phys. Chem. A*, **2001**, *105*, 293.
38) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
39) Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.
40) Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 306.
41) Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *103*, 379.
42) Amos, R. D.; Handy, N. C.; Knowles, P. J.; Rice, J. E.; Stone, A. J. *J. Phys. Chem.* **1985**, *89*, 2186.
43) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1977**, *66*, 1496.
44) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726.
45) Slipchenko, L. V.; Gordon, M. S. *Mol. Phys.* **2009**, *107*, 999.
46) Yamaguchi, Y.; Goddard, J. D.; Osamura, Y.; Schaefer, H. F. *A New Dimension in Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*; Oxford University Press: New York/Oxford, **1994**.
47) Jensen, J. H.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.
48) Kemp, D. D.; Rintelman, J. M.; Gordon, M. S.; Jensen, J. H. *Theor, Chem. Acc.* **2010**, *125*, 481.
49) Jensen, J. H. *J. Chem. Phys.* **1996**, *104*, 7795.
50) Li, H.; Gordon, M. S.; Jensen, J. H. *J. Chem. Phys.* **2006**, *124*, 214108.
51) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
52) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science*, **1983**, *220*, 671.
53) Cerny, V. *J. Optimization Theory and Appl.* **1985**, *45*, 41.

CHAPTER 2. THE ELECTRON AFFINITY OF Al_{13}: A CORRELATED ELECTRONIC STRUCTURE STUDY

A paper published in and reprinted with permission from
Journal of Physical Chemistry A **2011,** *115*, 899

Q. A. Smith and M. S. Gordon

Abstract

Neutral and anionic 13-atom aluminum clusters are studied with high-level, fully *ab initio* methods: second-order perturbation theory (MP2) and coupled cluster theory with singles, doubles, and perturbative triples (CCSD(T)). Energies and vibrational frequencies are reported for icosahedral and decahedral isomers, and are compared with density functional theory results. At the MP2 level of theory, with all of the basis sets employed, the icosahedral structure is energetically favored over the decahedral structure for both the neutral and anionic Al_{13} clusters. Hessian calculations imply that only the icosahedral structures are potential energy minima. The CCSD(T)/aug-cc-pVTZ adiabatic electron affinity of Al_{13} is found to be 3.57 eV, in excellent agreement with experiment.

Introduction

Al_{13}^- is a remarkably stable closed-shell cluster that has been referred to as a "superhalogen"[1]. Therefore, Al_{13}^- is of potential interest as an anion in ionic liquids. Experimental studies suggest that Al_{13}^- is highly symmetric[2] and cannot be etched by oxygen[3]. These properties are in agreement with its description as a "magic cluster" according to the jellium model[4], in which the nuclear geometry of the cluster is approximated as a spherical charge distribution interacting with delocalized valence electrons. This model predicts that 2, 8, 20, 40, etc. valence electrons will correspond to stable, closed-shell clusters. Al_{13}^- has 40 valence electrons.

The structure of Al_{13}- is widely accepted to be icosahedral[1,5], having two five-atom "rings" that are staggered with respect to one another. For the neutral Al_{13} cluster, previous density functional theory (DFT) studies[6-8] have predicted two possible isomers: a distorted icosahedron and a distorted decahedron (Figures 2.1-2.2). In the decahedral structure, the two five-atom "rings" are aligned. A study[6] using the BPW91 functional[9] and the LANL2DZ basis set with the Hay-Wadt relativistic effective core potential[10] predicts the neutral Al_{13} decahedron to be energetically preferred over the icosahedron by 0.43 eV. A study[7] using the hybrid B3LYP functional[11] with the same LANL2DZ likewise finds the neutral decahedron to be 0.41 eV lower in energy than the icosahedron. However, the energy ordering reverses when an all-electron calculation is performed[8]: the icosahedron is 0.23 eV lower in energy at the BPW91/6-311G(d) level of theory. The authors describe the isomers as nearly degenerate within the accuracy of the calculations[6,8].

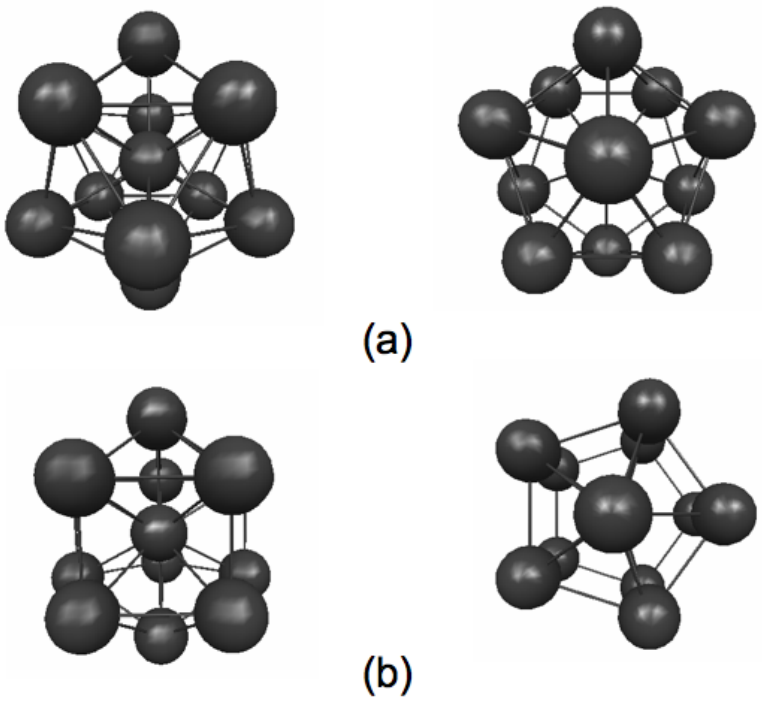


Figure 2.1. Side and top views of (a) the icosahedral and (b) the decahedral geometries of the 13-atom aluminum cluster. The anion takes on perfect (non-distorted) geometries while the neutral cluster geometries are Jahn-Teller distorted.

Using the local spin density approximation with the VWN functional[12], Mananes et al[13] conclude that the neutral Al_{13} cluster takes on a distorted icosahedral D_{3d} geometry. Han,

Jung, and Kim[14] show that B3LYP with the aug-cc-pVDZ basis set[15] over-distorts the neutral Al_{13} cluster, producing a heavily distorted C_s structure as the ground state. This B3LYP/aug-cc-pVDZ structure does not resemble either an icosahedron or a decahedron. At this level of theory, the D_{3d} geometry is 0.111 eV higher in energy than the C_s structure.[14] The hybrid PBE functional[16] with the aug-cc-pVDZ basis set, however, predicts a D_{3d} ground state, with the C_s structure 0.258 eV higher in energy.[17]

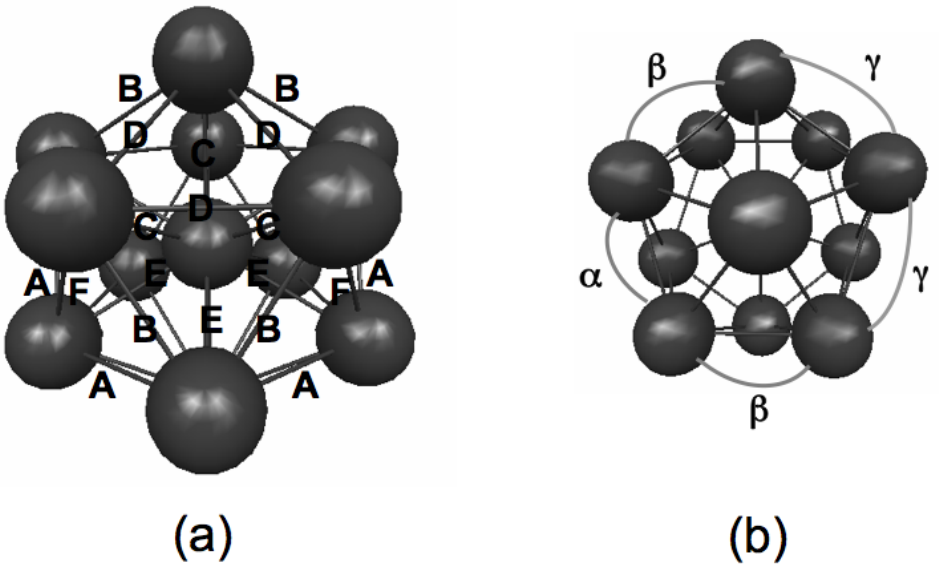


Figure 2.2. Bond (a) lengths and (b) angles of the Jahn-Teller distorted icosahedral structure of neutral Al_{13}. In (a), A=2.901 Å, B=2.772 Å, C=2.659 Å, D=2.971 Å, E=2.734 Å, F=2.764 Å. In (b), representative angles α=60.0°, β=57.6°, γ=63.2°. For the perfect icosahedral anion, all bond lengths to interior atom are 2.685 Å and between exterior atoms are 2.823 Å; all angles are 60.0°. Bond lengths and angles determined with MP2/6-311+G(2d).

Other studies describe the decahedral and icosahedral neutral isomers as competing structures due to their small energy difference.[18] A Car-Parrinello study[19] finds a single, well-defined, slightly distorted icosahedral minimum energy structure for Al_{13}, while a model potential basin-hopping Monte Carlo simulation[20] finds the point group of Al_{13} to be I_h. A study[21] utilizing a genetic algorithm coupled with a tight-binding potential, whose low-energy structure candidates were further optimized with local density approximation DFT, finds an icosahedral ground state for Al_{13}. In this study, the decahedral structure is 0.48 eV higher in energy, and the heavily distorted C_s structure discussed previously (found as the ground state with B3LYP/aug-cc-pVDZ in Ref. 14) lies about 0.02 eV higher still.[20]

Al_{13}- has been the subject of a coupled cluster study[21], in which the singles + doubles + non-iterative pertubative triples, CCSD(T)[22], method was employed. CCSD(T) single point energy calculations, using the 6-311G(d)[23] basis, were performed on the icosahedral and decahedral isomers of Al_{13}- after optimization with the Hartree-Fock (HF)/3-21G(d) method. At the CCSD(T)/6-311G(d) level of theory, the icosahedron is 0.54 eV lower in energy than the decahedron. Furthermore, the authors report that electron propagator[24] vertical electron detachment energy (VEDE) predictions for the icosahedron are in much closer agreement with experimental photoelectron spectra than are the predicted VEDEs for the decahedron.[21]

Distortion from the perfect icosahedral and perfect decahedral structures in the neutral isomers has been attributed to the Jahn-Teller effect.[6,8] However, a thorough molecular orbital investigation of this phenomenon has not been presented, to the authors' knowledge.

Explicit mention of vibrational frequencies or the outcome of Hessian (matrix of energy second derivatives) calculations for Al_{13} species is also rare in the literature, with the notable exception of Ref. 14 for $Al_{13}H$. Hessian matrices in particular are important, since a positive definite Hessian (all-real vibrational frequencies) indicates that one has found a local minimum on the potential energy surface. On the other hand, one or more negative Hessian eigenvalues (imaginary frequencies) indicate that the found structure is a saddle point. A single imaginary frequency suggests a transition state. Most of the works cited in the previous paragraphs do not provide such information.

Examining the adiabatic electron affinity of Al_{13} gives insight into the importance of geometric relaxation upon electron attachment, as well as an assessment of the accuracy of the geometry optimizations. The vertical electron affinity (VEA) is the energy difference between the neutral cluster and the anionic cluster at the geometry of the neutral cluster. On the other hand, the adiabatic electron affinity (AEA) is the energy difference between the neutral and anionic clusters at their respective optimized geometries. So, the AEA includes relaxation effects that occur after the electron attachment occurs. Similar values for the vertical and adiabatic electron affinities imply that the neutral and anionic clusters have similar geometries. Since experimental results for electron affinities may be available even when experimental structures are not, an energy difference between theoretically determined

structures that is in good agreement with experimental electron affinity values can imply the validity of the theoretical structures. A photoelectron spectroscopy study[2] found the adiabatic electron affinity of the neutral cluster to be 3.62 ± 0.06 eV. This provides impetus for the present computational analysis. In addition, while the structure of the anionic cluster Al_{13}^- appears to be well established, there is clearly little consensus on the structure of the neutral cluster, or on the relative energies of the two species. Therefore, an accurate determination of the structure of Al_{13} and the related energetics is important.

Methods

Second-order Moller-Plesset perturbation theory (MP2)[25] was employed as a means of studying the neutral and anionic Al_{13} clusters with a fully *ab initio* method that includes electron correlation. MP2 geometry optimizations and semi-numerical Hessians (finite differencing of analytic gradients) were obtained using three basis sets: LANL2DZ, which uses an effective core potential (ECP), and the all-electron Pople bases 6-31+G(d)[26] and 6-311+G(2d).[23] Additionally, single-point energy calculations on the MP2/6-311+G(2d) optimized icosahedral structures of anionic and neutral Al_{13} were carried out using coupled cluster theory with singles, doubles, and perturbative triples (CCSD(T)[22]) with the aug-cc-pVTZ basis[15]. Theoretical values for the adiabatic ionization potential of Al_{13}^- were determined by taking the energy difference between the anionic and neutral clusters at their respective optimized geometries.

For comparison, optimizations and semi-numerical Hessian calculations were also performed using the same three basis sets with the B3LYP and BPW91density functionals, using the geometries of Rao and Jena[6] as a starting point.

The CCSD(T) single-point energies were determined using Molpro[27]. All other calculations were performed with the GAMESS (General Atomic and Molecular Electronic Structure Theory) quantum chemistry software package[28].

Results and Discussion

At the MP2 level of theory, with any of the basis sets examined, the icosahedral structure is energetically favored over the decahedral structure for both the neutral and anionic Al_{13} clusters (Tables 2.1-2.2). Additionally, Hessian calculations imply that only the icosahedral structures are potential minima, having no imaginary frequencies (Tables 2.3-2.4).

Table 2.1. Energy difference (in eV) between the decahedral (D_{5h}) and icosahedral (D_{5d}) isomers of anionic Al_{13}. A positive value indicates that the icosahedral structure is lower in energy than the decahedral structure; a negative value indicates the decahedron is preferred.

	BPW91	B3LYP	MP2
LANL2DZ	-0.12	-0.36	0.50
6-31+G(d)	0.44	0.17	1.59
6-311+G(2d)	0.43	0.17	1.40

Table 2.2. Energy difference (in eV) between the distorted decahedral (C_s) and distorted icosahedral (D_{3d}) isomers of neutral Al_{13}. A positive value indicates the distorted icosahedral structure is lower in energy than the distorted decahedral structure; a negative value indicates the distorted decahedron is preferred.

	BPW91	B3LYP	MP2
LANL2DZ	-0.31	-0.53	0.04
6-31+G(d)	0.26	-0.01	1.02
6-311+G(2d)	0.26	0.03	0.92

First, consider the relative energies listed in Tables 2.1 and 2.2. MP2 predicts that the icosahedral Al_{13}^- structure is lower in energy than the decahedral structure with all basis sets, although the LANL2DZ effective core potential (ECP) predicts an energy difference that is too small. Indeed, this basis set is inadequate at all levels of theory. Both functionals, BPW91 and B3LYP, predict the incorrect energy order when the ECP basis set is used. Even with the larger, all-electron basis sets, DFT predicts an energy difference that is too small, even though the correct energy order is predicted. Similar comments apply to the neutral Al_{13} cluster. The ECP basis set predicts energy differences that are too small or the wrong sign, and the DFT methods with the all-electron basis sets predict the correct energy order with a magnitude that is too small.

MP2 consistently predicts that the icosahedral structure is a minimum on both the Al_{13} and Al_{13}^- potential energy surfaces (Table 2.3). For the anion, MP2 predicts the decahedral structure to be a first order saddle point (transition state). The two functionals are less definitive, since their prediction regarding the nature of this stationary point depends on the basis set that is used. For the largest basis set, all three methods agree that the decahedral species is a transition state. For the neutral Al_{13} radical, the decahedral species is clearly not a minimum on the potential energy surface. The number of imaginary frequencies depends on both the method and the basis set, and the ECP is clearly unreliable.

Table 2.3. Number of imaginary frequencies found for each isomer of the anionic Al_{13} cluster.

	Icosahedron (D_{5d})			Decahedron (D_{5h})		
	BPW91	**B3LYP**	**MP2**	**BPW91**	**B3LYP**	**MP2**
LANL2DZ	5	5	0	0	2	1
6-31+G(d)	0	0	0	1	0	N/A
6-311+G(2d)	0	0	0	1	1	1

Table 2.4. Number of imaginary frequencies found for each isomer of the neutral Al_{13} cluster.

	Distorted Icosahedron (D_{3d})			Distorted Decahedron (C_s)		
	BPW91	**B3LYP**	**MP2**	**BPW91**	**B3LYP**	**MP2**
LanL2DZ	4	6	0	0	0	0
6-31+G(d)	0	2	0	2	1	1
6-311+G(2d)	0	0	0	2	2	7

The icosahedral anion has perfect I_h symmetry (specified as D_{5d} in practice). At the MP2/6-311+G(2d) level of theory, the nearest neighbor bond distances are 2.823 Å between the outer twelve atoms and 2.685 Å from the outer atoms to the interior atom. This agrees nearly exactly with B3LYP/6-311+G(2d) (2.820 Å exterior, 2.682 Å interior) and very closely with BPW91/6-311+G(2d) (2.811 Å exterior, 2.673 Å interior). The decahedral anion with D_{5h} symmetry has one imaginary frequency [112i cm^{-1} with MP2/6-311+G(2d)], consisting of a torsional motion that would rotate the aligned pentagonal "rings" to a staggered configuration as in the icosahedron (Figure 2.3). As noted above, this suggests the decahedral structure is a transition state.

The neutral (doublet radical) structure is distorted from its closed shell anionic

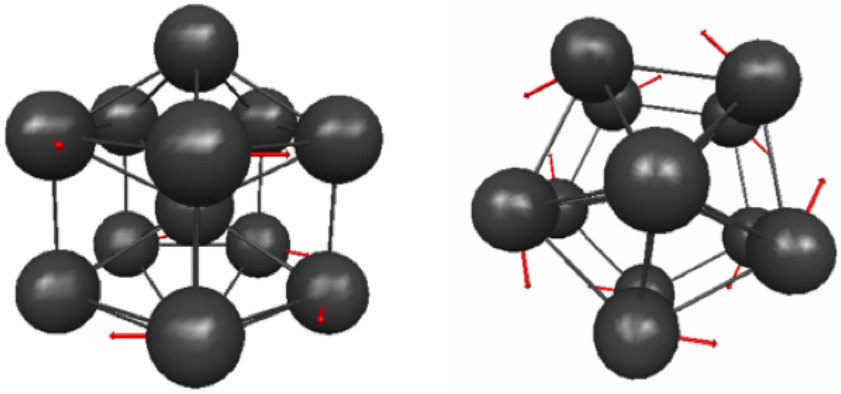


Figure 2.3. Side and top views of the imaginary frequency mode from the anionic cluster constrained to D_{5h} symmetry. This torsional mode suggests the structure will naturally take icosahedral symmetry.

analogue. When constrained to the same D_{5d} symmetry as the undistorted icosahedral anion, the neutral structure exhibits multiple imaginary frequencies (Table 2.4). Consistent with a previously proposed distorted icosahedral D_{3d} structure[13], the neutral cluster has a positive definite Hessian when relaxed to D_{3d} symmetry. At the MP2 level of theory, this distortion from D_{5d} to D_{3d} lowers the energy by 0.2 eV. Decahedral structures for the neutral cluster exhibit multiple imaginary frequencies, whether constrained to (undistorted) D_{5h} symmetry or relaxed to (distorted) C_s symmetry.

The MP2/6-311+G(2d) neutral icosahedral bond lengths range from 2.901 Å to 2.734 Å. The B3LYP/6-311+G(2d) and BPW91/6-311+G(2d) ranges of distances are 2.992 Å to 2.723 Å and 2.964 Å to 2.663 Å, respectively. Bond lengths as shown in Fig. 2.2 can be found in Table 2.5.

Table 2.5. Bond lengths (Å) in distorted icosahedral (D_{3d}) geometry of neutral Al_{13}. Letters A-F correspond to bonds shown in Fig. 2.2(a). Calculations performed with 6-311+G(2d) basis set.

	BPW91	B3LYP	MP2
A	2.879	2.898	2.901
B	2.762	2.763	2.772
C	2.663	2.677	2.659
D	2.964	2.992	2.971
E	2.712	2.723	2.734
F	2.763	2.780	2.764

The distortion in the neutral clusters arises from the Jahn-Teller effect.[6,8] This requires the unequal occupancy of two (or more) degenerate orbitals. Jahn-Teller distortion removes the degeneracy by lowering the symmetry. Orbital diagrams of the anionic (D_{5d} icosahedral) and neutral (D_{3d} distorted icosahedral) Al_{13} clusters appear in Fig. 2.4. The highest occupied molecular orbitals (HOMO) of the anion (Fig. 2.4(a)) are doubly degenerate orbitals with E_{1u} symmetry. They are isoenergetic with a second doubly degenerate set of orbitals having E_{2u} symmetry. Removing an electron from the E_{1u} HOMO to produce neutral Al_{13} yields the unequally occupied degenerate orbitals that lead to Jahn-Teller distortion. The distortion of the neutral cluster produces a singly occupied, nondegenerate HOMO with A_{2u} symmetry (Fig. 2.4(b)). The Al_{13} HOMO is nondegenerate A_{2u} even when the neutral cluster is constrained to the same (nondistorted) icosahedral geometry as the anion.

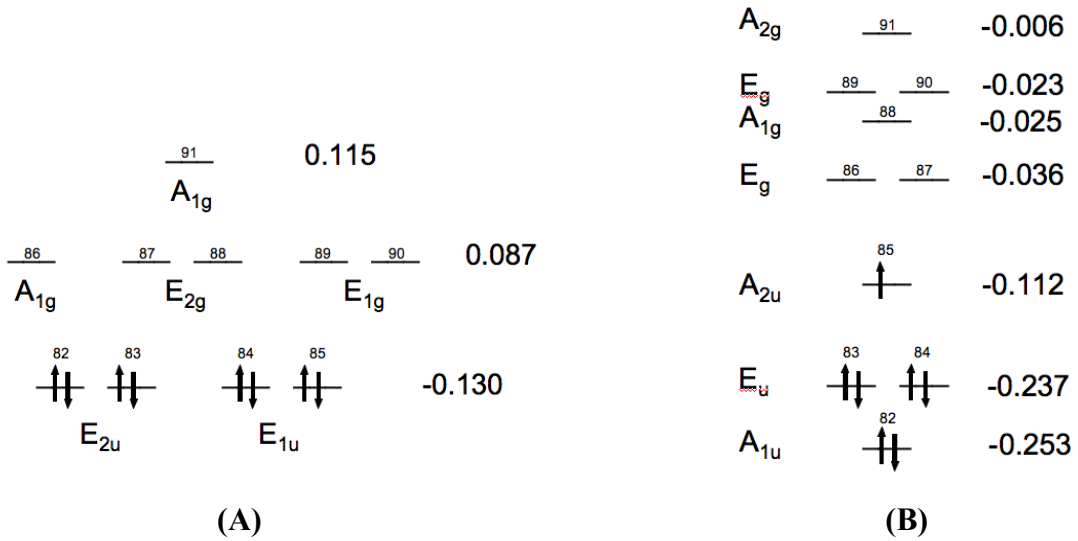


Figure 2.4. Orbital diagram of (a) anionic and (b) neutral Al_{13} clusters. Orbital symmetries and energies (in hartree) shown for a selection of orbitals near the HOMO-LUMO. The anion has perfect icosahedral symmetry (D_{5d} in practice), while the neutral has distorted icosahedral symmetry (D_{3d}).

The adiabatic electron affinity of Al_{13} is the energy difference between the optimized neutral and anionic structures (Table 2.6). Relative to the experimental value of 3.62 ± 0.06 eV, MP2 overestimates the adiabatic electron affinity by about 0.2 eV when either Pople basis set is used, but underestimates it by 0.15 eV with the LANL2DZ basis. DFT calculations underestimate this value to an extent that is dependent on the functional. With

BPW91, the calculated adiabatic electron affinity is about 0.3 eV lower than the experimental value. B3LYP yields a more significant deviation, approximately 0.5 eV lower. Single point energy calculations using CCSD(T)/aug-cc-pVTZ on the geometries obtained with MP2/6-311+G(2d) optimizations yields an adiabatic electron affinity of 3.57 eV, in excellent agreement with the experimental value.

Table 2.6. Adiabatic electron affinity of Al_{13}. Units of eV. These values are the energy difference between the neutral icosahedral (D_{3d}) and anion icosahedral (D_{5d}) structures. Experimental value is 3.62 ± 0.06 eV (Ref. 2).

	BPW91	B3LYP	MP2	CCSD(T)
LanL2DZ	3.28	3.06	3.47	
6-31+G(d)	3.34	3.16	3.83	
6-311+G(2d)	3.33	3.14	3.86	3.57

The MP2/6-311+G(2d) vertical ionization potential (the energy difference between the optimized neutral cluster and the cation at the neutral cluster geometry) is 6.93 eV for the decahedron and 7.21 eV for the icosahedron. DFT[8] with the BPW91/6-311G(d) functional predicts a vertical ionization potential for the decahedron to be 6.48 eV and that of the icosahedron to be 6.79 eV. An experimental study[29] reports the vertical electron affinity to be 6.45±0.05 eV. So, there is a very large geometry relaxation effect on the observed electron affinity.

Conclusions

A small, frozen core basis set like LANL2DZ is inadequate to describe the electronic structure of Al_{13} clusters, especially when used with density functional theory calculations. Compared to results found with the larger, all-electron Pople basis 6-311+G(2d), DFT calculations with LANL2DZ show a reversal of the energetic ordering of the icosahedral and decahedral isomers of both neutral and anionic Al_{13}.

DFT predictions vary with the choice of functional, notably in the case of the ground state neutral Al_{13} structure. B3LYP with either all-electron basis assigns very nearly equal energies to the decahedral and icosahedral geometries (differing by hundredths of an eV), while BPW91 favors the icosahedron by 0.26 eV. MP2 predicts that the neutral icosahedron

is approximately 1 eV more stable than the decahedron. B3LYP also underestimates the energy difference between isomers of the anionic cluster, giving 0.17 eV versus 0.43 eV for BPW91/6-311+G(2d) and 1.40 for MP2/6-311+G(2d). However, icosahedral bond lengths and angles determined with B3LYP, rather than BPW91, more closely match those determined with MP2. Hessian calculations, which can show whether the proposed structure is indeed a local minimum on the potential energy surface, are also dependent on the choice of functional.

The Al_{13} adiabatic electron affinity value of 3.57 eV found via CCSD(T)/aug-cc-pVTZ single point energy calculations on MP2/6-311+G(2d) geometries is in excellent agreement with the experimentally determined value[2] of 3.62 ± 0.06 eV.

Acknowledgements

This work was supported by a grant from the Air Force Office of Scientific Research. Helpful discussions with Professor Will Castleman are gratefully acknowledged.

References

1) Bergeron, D. E.; Roach, P. J.; Castleman, A. W., Jr.; Jones, N. O.; Khanna, S. N. *Science* **2004**, *304*, 84
2) Li, X.; Wu, H.; Wang, X. B.; Wang, L. S. *Phys. Rev. Lett.* **1998**, *81*, 1090
3) Leuchtner, R. E.; Harms, A. C.; Castleman, A. W., Jr. *J. Chem. Phys.* **1989**, *91*, 2753
4) Knight, W. G.; Clemenger, K.; de Heer, W. A.; Saunders, W. A.; Chou, M. Y.; Cohen, M. L. *Phys. Rev. Lett*. **1984**, *52*, 2141
5) Khanna, S. N.; Rao, B. K.; Jena, P. *Phys. Rev. B* **2002**, *65*, 125105
Zheng, W. J.; Thomas, O. C.; Lippa, T. P.; Xu, S. J.; Bowen, K. H., Jr. *J. Chem. Phys.* **2006**, *124*, 144304
6) Rao, B. K.; Jena, P. *J. Chem. Phys.* **1990**, *111*, 1890
7) Zope, R. R.; Baruah, T. *Phys. Rev. A* **2001**, *64*, 053202
8) Rao, B. K.; Khanna, S. N.; Jena, P. *Phys. Rev. B* **2000**, *62*, 4666
9) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098
Perdew, J. P. *Electronic Structure of Solids '91*, Eds. P. Ziesche and H. Eschrig; Akademie Verlag: Berlin, **1991**, p. 11
10) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284
11) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648
Lee, C.; Yang, W.; Parr, R.G. *Phys. Rev. B* **1988**, *37*, 785

Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623
12) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys*. **1980**, *58*, 1200
13) Mañanes, A.; Duque, F.; Méndez, F.; López, M. J.; Alonso, J. A. *J. Chem. Phys*. **2003**, *119*, 5128
14) Han, Y.-K.; Jung, J.; Kim, K. H. *J. Chem. Phys*. **2005**, *122*, 124319
15) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358
16) Adamo, C.; Barone, V. *J. Chem. Phys*. **1999**, *110*, 6158
17) Akola, J.; Häkkinen, H.; Manninen, M. *Phys. Rev. B* **1998**, *58*, 3601
18) Yi, J. Y.; Oh, D. H.; Bernholc, J.; Car, R. *Chem. Phys. Lett*. **1990**, *174*, 461
19) Sebetci, A.; Güvenç, Z. B. *Modelling Simul. Mater. Sci. Eng*. **2005**, *13*, 683
20) Chuang, F. C.; Wang, C. Z.; Ho, K. H. *Phys. Rev. B* **2006**, *73*, 125431
21) Dolgounitcheva, O.; Zakrzewski, V. G.; Ortiz, J. V. *J. Chem. Phys*. **1999**, *111*, 10762
22) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910
Ragavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett*. **1989**, *157*, 479
23) McLean, A. D.; Chandler, G. S. *J. Chem. Phys*. **1980**, *72*, 5639
24) Ortiz, J. V. *Computational Chemistry: Reviews of Current Trends*, *Vol. 2*, Ed. J. Leszczynski; World Scientific: Singapore, **1997**, p. 1.
Ortiz, J. V.; Zakrzewski, V. G.; Dolgounitcheva, O. *Conceptual Perspectives in Quantum Chemistry, Vol. 3*, Eds. J.-L. Calais and E. Kryachko; Kluwer: Dordrecht, **1997**, p. 465
Ortiz, J. V. *Adv. Quantum Chem.* **1999**, *35*, 33
25) Møller, C.; Plesset, M. S. *Phys. Rev*. **1934**, *46*, 618
26) Francl, M. M.; Petro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys*. **1982**, *77*, 3654
27) MOLPRO, version 2006.1, a package of ab initio programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, and others, see http://www.molpro.net.
28) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347
29) Schriver, K. E.; Persson, R. J.; Honea, E. C.; Whetten, R. L. *Phys. Rev. Lett*. **1990**, *64*, 2539

CHAPTER 3. MODELING π-π INTERACTIONS WITH THE EFFECTIVE FRAGMENT POTENTIAL METHOD: THE BENZENE DIMER AND SUBSTITUENTS

A paper published in and reprinted with permission from
Journal of Physical Chemistry A **2008**, *112*, 5286

T. Smith, L. V. Slipchenko, and M. S. Gordon

Abstract

This study compares the results of the general effective fragment potential (EFP2) method to the results of a previous combined coupled cluster with single, double and perturbative triple excitations [CCSD(T)] and symmetry adapted perturbation theory (SAPT) study[1] on substituent effects in π-π interactions. EFP2 is found to accurately model the binding energies of the benzene-benzene, benzene-phenol, benzene-toluene, benzene-fluorobenzene, and benzene-benzonitrile dimers as compared with high level methods[1], but at a fraction of the computational cost of CCSD(T). Additionally, an EFP-based Monte Carlo/simulated annealing study was undertaken to examine the potential energy surface of the substituted dimers.

Introduction

Intermolecular π-π interactions are among the major noncovalent forces controlling structural organization and recognition processes in biomolecules.[2] Interactions between aromatic rings are largely responsible for DNA base-pair stacking,[3] host-guest complexation,[4-7] and the tertiary structure of proteins.[8,9] Certain drugs rely on π-π interactions for intercalation into DNA.[10,11] While these interactions have been studied extensively,[12-24] their relative weakness and shallow potential energy surface makes them challenging to describe by either experiment or theory.[25-29] The binding energy of the gas

phase benzene dimer, for example, is 2 - 3 kcal mol^{-1}, and the dimer is stable only at low temperatures.[30] Only the highest levels of electronic structure theory can accurately capture these weak interactions.

The fundamental nature of π-π interactions, along with the difficulties that theory encounters when targeting the multifaceted patterns of intermolecular bonding, has made the benzene dimer and its substituents popular benchmark systems for the accuracy and feasibility of new computational techniques and approaches.[1,30-39] Both experimental[18,28,40] and theoretical[30-35] studies suggest that the benzene dimer has two nearly isoenergetic geometries: T-shaped and parallel displaced. At the coupled cluster with singles, doubles, and perturbative triples [CCSD(T)][41] level in an estimated complete basis set limit, the gas-phase binding energies D_e (D_o) of these geometries were calculated to be 2.7 (2.4) and 2.8 (2.7) kcal mol^{-1}, respectively.[30] A sandwich geometry (face-to-face stacking) was found to be approximately 1 kcal mol^{-1} higher in energy and is a transition state.[30]

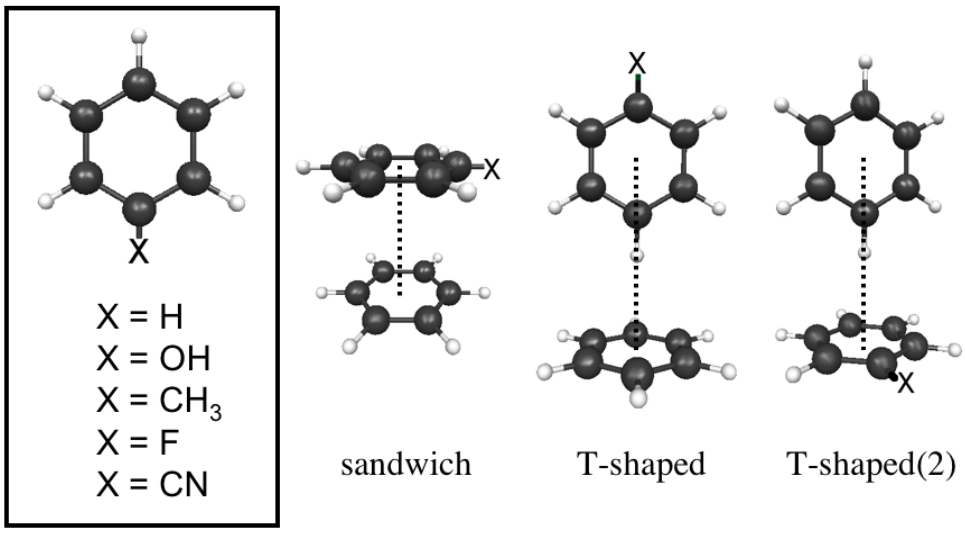


Figure 3.1. Substituted benzene dimer geometries. The geometry referred to as "T-shaped" in Ref. 1 has been renamed "T-shaped(1)".

Adding a substituent (OH, CH_3, F, or CN) to one of the benzene rings in the dimer can result in interesting changes to the π-π interaction energies.[1,37] Sherrill and coworkers studied the sandwich geometry and two different T-shaped geometries for several substituted benzene dimers (Figure 3.1). An estimated CCSD(T)/aug-cc-pVTZ level of theory predicts

that all substituted sandwich dimers have a larger binding energy than the unsubstituted benzene dimer. The sandwich configuration exhibits a larger stabilizing effect due to substituents than either of the T-shaped configurations does, so the energy difference between these configurations is smaller for the substituted dimers than it is for the unsubstituted benzene dimer. For the T-shaped(1) structures (referred to as "T-shaped" in ref. 1), the presence of an electron-withdrawing substituent (F, CN) results in a stronger interaction compared to the unsubstituted benzene dimer, while electron-donating substituents (OH, CH_3) interact more weakly than the unsubstituted dimer does. For the T-shaped(2) dimers, the opposite is true. It is noteworthy that, as discussed in ref. 1, the binding patterns found for the substituted benzene dimers contradict the Hunter-Sanders model[42], which qualitatively predicts binding energies based on electrostatics alone. This indicates the importance of the non-electrostatic terms in determining the binding energies of these dimers.

The same set of substituted benzene dimers was recently studied by using the van der Waals density functional theory (vdW-DFT).[36] Most DFT methods do not accurately describe dispersion interactions, and generalized gradient approximation (GGA) functionals often fail to predict any binding in molecular dimers.[36] However, the vdW-DFT method[43], which incorporates a nonlocal correlation energy, has been shown[36] to reproduce the energetic ordering of the substituents consistent with the CCSD(T) results of ref. 1. But, the interaction energies of the sandwich dimers are overestimated by 0.6-0.7 kcal mol^{-1} compared to CCSD(T), whereas the binding in the T-shaped dimers is underestimated by 0.2-0.4 kcal mol^{-1}.

In this contribution, the interactions in mixed substituted benzene dimers are studied by means of the general effective fragment potential (EFP2) method.[44,45] A benchmark study is presented, comparing the performance of EFP2 with the CCSD(T) and symmetry adapted perturbation theory (SAPT)[46] results of Sinnokrot and Sherrill.[1] EFP2 is an *ab initio*-based model potential method for treating intermolecular interactions, wherein parameters for all major types of non-covalent forces – Coulomb, exchange-repulsion, polarization (induction), and dispersion – are obtained for each unique fragment from a single *ab initio* calculation. Contrary to many other model potential methods, EFP2 does not use fitted parameters to

reproduce *ab initio* or experimental results, but rather derives the parameters directly from first principles. These parameters can then be used to model interactions with other EFP fragments or with fully *ab initio* molecules. EFP fragments have frozen internal geometries.

The computational cost of EFP2 is orders of magnitude lower than that of *ab initio* techniques. For example, for the benzene dimer in the 6-311+G(3df,2p) basis set with 660 basis functions, a single-point energy calculation with second order Moller-Plesset perturbation theory (MP2)[47] would require 142 minutes of CPU time on one IBM Power5 processor, whereas the analogous EFP calculation takes only 0.4 seconds[48]. If EFP2 can be shown to produce results of sufficient accuracy, the advantage of using it in place of more computationally costly methods, like CCSD(T) or MP2, is clear.

The effective fragment potential method has been shown to successfully model numerous hydrogen bonded systems. For example, EFP1 (the EFP-type model designed specifically for water, with a fitted exchange-repulsion term) has been used to model chemical reactions in solution,[49,50] solvent clusters,[51,52] an S_N2 reaction,[53] and amino acid neutral/zwitterion equilibria.[54,55] Styrene-styrene interactions[56] and methanol-water mixing[57] have been studied by the general EFP2 method. Recently, EFP2 was shown to successfully model the π-π interactions in the unsubstituted benzene dimer.[48] An EFP2 study of benzene-water complexes is in progress.[58]

The main goal of the present work is to further investigate the performance of EFP2 for systems with π-π interactions in order to confidently model various biological systems, e.g., interactions in DNA base pairs, in future studies. Therefore, the first part of this work analyzes the accuracy of EFP2 on a set of mixed benzene-substituted benzene dimers, comparing the EFP2 results with those of CCSD(T) and SAPT from ref. 1. The second part of this paper undertakes an independent EFP2 study of the dimer potential energy surfaces and bonding patterns.

This work is organized as follows. Section 2 describes the theoretical methods and computational details. Section 3 presents and discusses the results. The main conclusions are given in Section 4.

Methods

The EFP2 method has been implemented in the quantum chemistry program GAMESS[59], which was used for all calculations in this study.

Dimer geometries. To benchmark the performance of EFP2 versus CCSD(T) and SAPT, the monomer and dimer geometries were adopted from ref. 1. The procedure used in ref. 1 to find the dimer structures was as follows. The geometries of all monomers were optimized with MP2 and the aug-cc-pVTZ basis set.[60] For benzene, a C-C bond length of 1.3942 Å and C-H bond length of 1.0823 Å were obtained. The geometry of the toluene monomer was fixed at C_s symmetry, with one methyl hydrogen above the benzene ring and two below it, and phenol was chosen to be planar. In all dimer calculations in ref. 1, the monomers were held rigid, having fixed internal coordinates. The dimer geometries were obtained by moving the rigid monomers relative to one another in order to find the optimal intermonomer distance (defined as the distance between ring centers) for the sandwich, T-shaped(1), and T-shaped(2) configurations (see Fig. 3.1). These structures are not actual energy minima for the dimers in question, but are constrained model structures chosen in ref. 1 for the purpose of assessing the influence of substituents on the π-π interactions. For example, to minimize direct interactions between the substituent on one ring and the other unsubstituted ring, the substituent in the T-shaped dimers is held in a para position relative to the un-substituted benzene. This serves to maintain focus on the effect of the substituent on the π-π interactions.

In this study, the monomer geometries from ref. 1 were used to generate EFP2 interaction parameters with the 6-311++G(3df,2p) basis set. The underlying 6-311G basis set is accurate for modeling electrostatic interactions, while polarization functions are important for modeling dispersion interactions[57] and diffuse functions are required for accurate exchange-repulsion interactions[45]. Distributed multipoles on atoms and bond mid-points were generated using a numerical integration scheme.[48] The electrostatic charge-charge, charge-dipole, charge-quadrupole, and dipole-dipole energies were screened by charge-penetration damping functions, as described in ref. 48.

As in ref. 1, the intermonomer distances in each of the chosen dimer configurations (Fig. 3.1) were varied to find the optimal EFP2 dimer geometry. Performing the geometry optimizations in this way facilitates a direct comparison of the optimal intermonomer distances (R) and energies (E_{int}) with the corresponding CCSD(T) values from ref. 1.

Energy component comparison. In order to analyze the accuracy of each component (Coulomb, exchange-repulsion, polarization, and dispersion) of the EFP2 energy, separate EFP2 calculations were performed at the dimer geometries used in the SAPT/aug-cc-pVDZ calculations in ref. 1. Sandwich dimers have a 3.70 Å intermonomer separation, while a 4.90 Å separation was used for the T-shaped(1) and T-shaped(2) dimers.

Investigation of the EFP2 potential energy surface. In addition to direct comparisons of the accuracy of EFP2 to that of CCSD(T), MP2, and SAPT, a Monte Carlo/simulated annealing (MC/SA)[61] study was performed on the EFP2 dimers to investigate the potential energy surface of each dimer. To increase the conformational sampling, two temperature ranges were used in each case: 20,000 – 100 K and 3000 – 100 K. A sandwich geometry was used as the starting structure for each dimer. Geometry optimizations were performed every 10 steps.

Results and Discussion

Intermonomer separations and binding energies. Table 3.1 compares the optimal EFP2 distances and binding energies of the substituted dimers with the distances and energies found with the MP2 and CCSD(T) methods, as given in ref. 1 for the constrained sandwich, T-shaped(1), and T-shaped(2) structures. The optimal EFP2 intermonomer distances are consistently larger than the estimated CCSD(T)/aug-cc-pVTZ distances by 0.05 to 0.10 Å for the sandwich dimers, 0.15 to 0.25 Å for the T-shaped(1) dimers, and 0.15 to 0.21 Å for the T-shaped(2) dimers. Compared to CCSD(T), MP2 with the aug-cc-pVTZ basis set consistently underestimates the intermonomer distances by 0.1 to 0.2 Å for all dimer geometries. MP2/aug-cc-pVDZ results are in better agreement with those of CCSD(T), underestimating the intermonomer distances by at most 0.1 Å. This suggests that there is a fortuitous cancellation of errors for MP2 with the smaller basis set.

Table 3.1. Optimized intermonomer distances (R), in Ångstroms, and interaction energies (E_{int}), in kcal/mol, for the sandwich, T-shaped(1), and T-shaped(2) structures of the benzene-substituted dimers. MP2 and estimated CCSD(T) results from ref. 1.

X	method	Sandwich		T-shaped(1)		T-shaped(2)	
		R	E_{int}	R	E_{int}	R	E_{int}
H	MP2/aug-cc-pVDZ	3.80	-2.90	5.01	-3.16	5.01	-3.16
	MP2/aug-cc-pVTZ	3.70	-3.26	4.89	-3.46	4.89	-3.46
	CCSD(T)/aug-cc-pVTZ	3.90	-1.80	4.99	-2.62	4.99	-2.62
	EFP2	3.95	-2.17	5.15	-2.42	5.15	-2.42
OH	MP2/aug-cc-pVDZ	3.70	-3.40	5.00	-3.14	4.95	-3.23
	MP2/aug-cc-pVTZ	3.60	-3.75	4.90	-3.42	4.90	-3.52
	CCSD(T)/aug-cc-pVTZ	3.80	-2.17	5.00	-2.58	5.00	-2.67
	EFP2	3.90	-2.72	5.15	-2.54	5.15	-2.45
CH_3	MP2/aug-cc-pVDZ	3.70	-3.58	5.00	-3.11	4.90	-3.60
	MP2/aug-cc-pVTZ	3.65	-3.96	4.90	-3.39	4.80	-3.89
	CCSD(T)/aug-cc-pVTZ	3.80	-2.27	5.00	-2.55	5.00	-2.95
	EFP2	3.90	-2.78	5.20	-2.47	5.15	-2.95
F	MP2/aug-cc-pVDZ	3.70	-3.50	4.95	-3.35	5.00	-2.87
	MP2/aug-cc-pVTZ	3.70	-3.81	4.90	-3.61	4.90	-3.17
	CCSD(T)/aug-cc-pVTZ	3.80	-2.29	5.00	-2.77	5.00	-2.38
	EFP2	3.90	-3.02	5.15	-2.79	5.20	-2.30
CN	MP2/aug-cc-pVDZ	3.70	-4.49	4.90	-3.79	5.00	-2.82
	MP2/aug-cc-pVTZ	3.60	-4.86	4.80	-4.11	4.90	-3.08
	CCSD(T)/aug-cc-pVTZ	3.80	-3.05	4.90	-3.25	5.00	-2.20
	EFP2	3.85	-3.91	5.15	-3.20	5.15	-2.23

For the sandwich dimers, EFP2 overestimates the magnitude of the binding energies by 0.36 to 0.86 kcal mol^{-1} compared to CCSD(T). In all cases, this represents a significant improvement over MP2, which overestimates the binding energy by 1.1 to 1.8 kcal mol^{-1} compared with the CCSD(T) values. EFP2 most accurately models the unsubstituted benzene sandwich dimer, giving a 0.37 kcal mol^{-1} error in the binding energy as compared to the estimated CCSD(T) energy. The benzene-benzonitrile sandwich dimer binding energy has the highest error at 0.86 kcal mol^{-1}. Compared to CCSD(T), MP2 underestimates the intermonomer separation and overestimates the binding energies of the dimer, whereas EFP2 overestimates both the intermonomer distances and binding energies. The overestimation of binding energies occurs to a much smaller degree with EFP2 than with MP2. A disconcerting observation is that the MP2 error in the binding energy increases when the basis set is improved.

EFP2 is very accurate in calculating the binding energies of the T-shaped(1) and T-shaped(2) dimers. The EFP2 binding energy of the T-shaped(1) unsubstituted benzene dimer (identical to the T-shaped(2) unsubstituted dimer) is higher (less strongly bound) than the corresponding CCSD(T) energy by 0.20 kcal mol^{-1}. Of the substituted T-shaped(2) dimers, benzene-phenol is the most in error at 0.22 kcal mol^{-1} higher in energy. Discrepancies in the other EFP2 T-shaped dimers range from 0.02 kcal mol^{-1} lower in energy (more strongly bound) to 0.08 kcal mol^{-1} higher in energy compared to CCSD(T). The EFP2 benzene-toluene energy is in perfect agreement with the estimated CCSD(T) energy for that dimer. Contrary to the excellent performance of EFP2, MP2 overestimates the binding energies of T-shaped(1) and T-shaped(2) dimers, becoming even less accurate when increasing the basis set from aug-cc-pVDZ to aug-cc-pVTZ.

These EFP2 results also compare favorably with the results of a DFT study[36] performed on the same set of substituted benzene dimers using a van der Waals (vdW)-corrected density functional.[43] This vdW-DFT method overestimates the magnitude of the sandwich dimer interaction energies by 0.56 to 0.68 kcal mol^{-1} compared to CCSD(T) and underestimates the binding in the T-shaped dimers, by 0.28 to 0.39 kcal mol^{-1} in the T-shaped(1) and by 0.19 to 0.34 kcal mol^{-1} in the T-shaped(2) dimers. In comparison, EFP2 overbinds the sandwich dimers by 0.36 to 0.86 kcal mol^{-1} compared to CCSD(T) and differs from the CCSD(T) results for the T-shaped dimers by less than 0.1 kcal mol^{-1} (except for the unsubstituted dimer and the T-shaped(2) benzene-phenol dimer, which are 0.2 kcal mol^{-1} less strongly bound). EFP2 requires considerably less computer time (on the order of seconds) once the MAKEFP potentials have been generated. In contrast, the GGA portion of the vdW-DFT calculations requires 1.3 h on a single Opteron processor.[36]

Energy component comparison. Table 3.2 and Figs. 3.2-3.4 compare the SAPT (from ref. 1) and EFP2 Coulomb, exchange-repulsion, polarization (or induction), and dispersion energies for the three types of constrained geometries. The energy component comparison was carried out at fixed intermonomer distances (3.70 Å for the sandwich structures, 4.90 Å for both types of T-shaped structures) that were smaller than the EFP2 equilibrium intermonomer distances but close to the estimated CCSD(T) equilibrium distances (see Table 3.1).

Table 3.2. Contributions to the binding energy (kcal/mol) for the sandwich, T-shaped(1), and T-shaped(2) dimers. The intermonomer separations are 3.70 Angstrom in the sandwich dimers and 4.90 Angstrom in the T-shaped(1) and T-shaped(2) dimers. SAPT results are from ref. 1.

X	Energy Component	Sandwich		T-shaped(1)		T-shaped(2)	
		SAPT	EFP2	SAPT	EFP2	SAPT	EFP2
H	Coulomb	-0.97	-0.70	-2.24	-2.71	-2.24	-2.71
	ex.-repulsion	6.03	6.03	4.87	4.96	4.87	4.96
	polarization	-0.33	-0.41	-0.67	-0.31	-0.67	-0.31
	dispersion	-6.53	-6.38	-4.37	-3.83	-4.37	-3.83
	total energy	-1.80	-1.46	-2.42	-1.89	-2.42	-1.89
OH	Coulomb	-1.08	-1.03	-2.21	-2.69	-2.18	-2.64
	ex.-repulsion	5.85	5.75	4.85	5.05	4.78	4.86
	polarization	-0.34	-0.39	-0.66	-0.33	-0.66	-0.27
	dispersion	-6.72	-6.55	-4.37	-4.07	-4.41	-3.86
	total energy	-2.29	-2.22	-2.39	-2.04	-2.47	-1.91
CH_3	Coulomb	-1.03	-0.65	-2.24	-2.63	-2.38	-2.69
	ex.-repulsion	6.21	5.85	5.02	5.24	4.81	4.76
	polarization	-0.40	-0.42	-0.67	-0.34	-0.70	-0.35
	dispersion	-7.19	-7.03	-4.46	-4.16	-4.59	-4.26
	total energy	-2.41	-2.26	-2.34	-1.89	-2.85	-2.54
F	Coulomb	-1.36	-1.54	-2.27	-2.82	-1.98	-2.43
	ex.-repulsion	5.73	5.63	4.55	4.72	4.73	4.82
	polarization	-0.29	-0.31	-0.68	-0.37	-0.57	-0.20
	dispersion	-6.49	-6.31	-4.22	-3.92	-4.30	-3.92
	total energy	-2.40	-2.53	-2.63	-2.39	-2.12	-1.73
CN	Coulomb	-1.83	-1.92	-2.59	-3.22	-1.73	-1.96
	ex.-repulsion	5.78	5.55	4.59	4.77	4.69	4.49
	polarization	-0.29	-0.29	-0.83	-0.53	-0.51	-0.10
	dispersion	-7.01	-6.87	-4.29	-3.83	-4.53	-4.20
	total energy	-3.36	-3.54	-3.12	-2.81	-2.09	-1.77

As described in ref. 1, SAPT predicts that dispersion is the single greatest attractive contribution to the overall binding energy of the dimers, especially for the sandwich structures. The sandwich dimers with electron-withdrawing substituents (CN and F) have the most favorable Coulomb interaction energy coupled with the lowest exchange-repulsion energy, giving them the strongest sandwich binding energies. The order of the total binding energies for the other sandwich dimers corresponds to the order of their dispersion energies. Due to the greater intermonomer separation in the T-shaped dimers, the exchange-repulsion and dispersion components of their binding energies are smaller than those of the sandwich structures. The Coulomb attractions in the T-shaped(1) and T-shaped(2) dimers are stronger

compared to the sandwich structures because the negatively charged π cloud of one monomer interacts with the positively charged hydrogens of the other.

Overall, good agreement is found between the EFP and SAPT energy terms (Table 3.2, Figs. 3.2-3.4). The EFP2 dispersion interaction is consistently lower in magnitude than that predicted by SAPT, but always by less than 0.5 kcal mol^{-1}; dispersion remains the dominant attractive contribution to the binding energy for both the sandwich and T-shaped dimers. The EFP2 polarization terms for the sandwich dimers agree to within ~0.1 kcal mol^{-1} with those predicted by SAPT. For the T-shaped(1) and T-shaped(2) dimers, the polarization terms are smaller in magnitude by 0.3 - 0.4 kcal mol^{-1} compared with the SAPT results. The EFP2 Coulomb term is slightly more negative than the corresponding SAPT term in the sandwich dimers with electron-withdrawing substituents (CN, F), but less negative in the benzene-benzene and benzene-toluene dimers. The EFP2 Coulomb interaction for the T-shaped structures is ~0.5 kcal mol^{-1} lower in energy on average compared to SAPT. The EFP2 exchange-repulsion generally differs from that obtained from SAPT by less than 0.1 kcal mol^{-1}.

For the sandwich structures (Figure 3.2A,B), the SAPT binding energy increases in magnitude in the order H < OH < F ≈ CH_3 < CN; the order predicted by both EFP2 and CCSD(T) is a very similar H < OH < CH_3 < F < CN (see Table 3.1). Overall, the trends predicted by EFP2 for the sandwich dimers agree with those found with SAPT in Ref. 1.

In the first set of T-shaped dimer structures (Figure 3.3A,B), the magnitudes of the dispersion and the exchange-repulsion terms are significantly reduced compared to their values in the sandwich dimer structures. This is reflected in both the SAPT and the EFP2 results. The order of increasing strength of the total interaction energy is CH_3 ≈ OH ≈ H < F < CN and CH_3 ≈ H < OH < F < CN for SAPT and EFP2, respectively. Taking into account the very small (< 0.1 kcal mol^{-1}) binding energy differences between CH_3, OH, and H-substituted dimers, the agreement between SAPT and EFP2 is very reasonable. The overall EFP2 binding energies are 0.2 to 0.5 kcal mol^{-1} lower in magnitude than SAPT binding energies. This reflects the fact that the chosen intermonomer separations are shorter than the optimal EFP2 geometries.

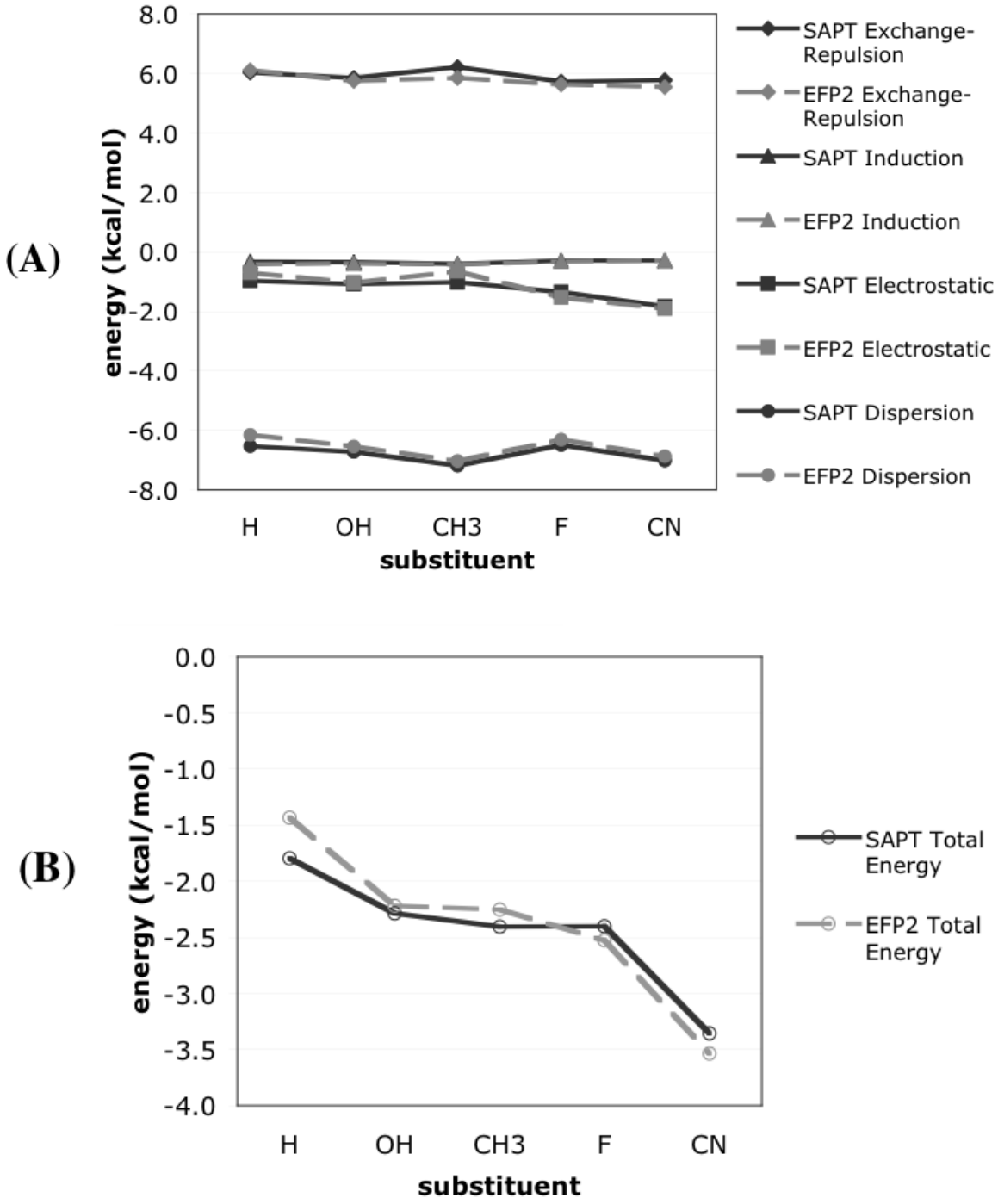


Figure 3.2. (A) Electrostatic, exchange-repulsion, polarization (induction), and dispersion components and **(B)** total interaction energies for the sandwich dimers computed with EFP2 and SAPT. SAPT data are from ref. 1.

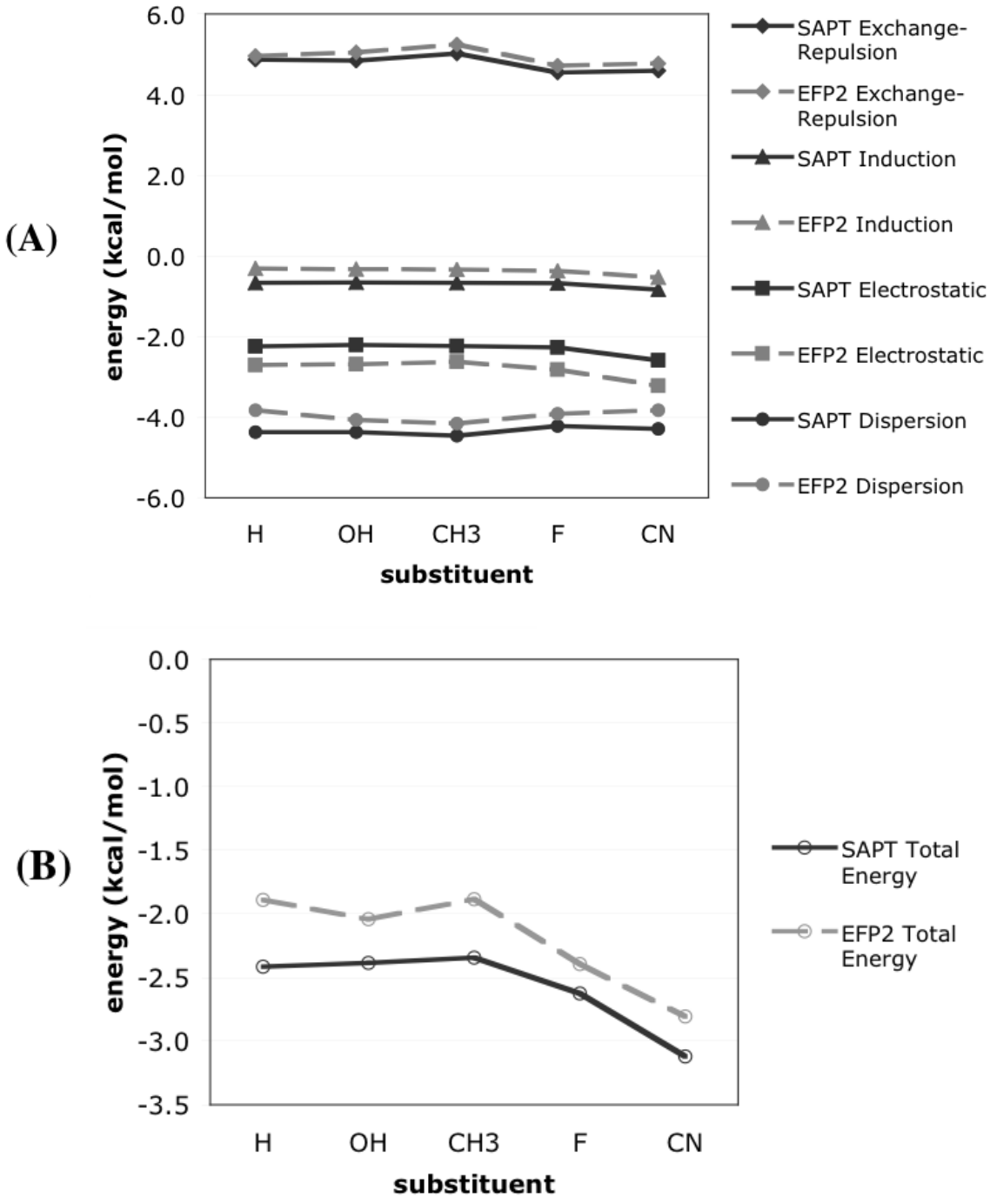


Figure 3.3. (A) Electrostatic, exchange-repulsion, polarization (induction), and dispersion components and **(B)** total interaction energies for the T-shaped(1) dimers computed with EFP2 and SAPT. SAPT data are from ref. 1.

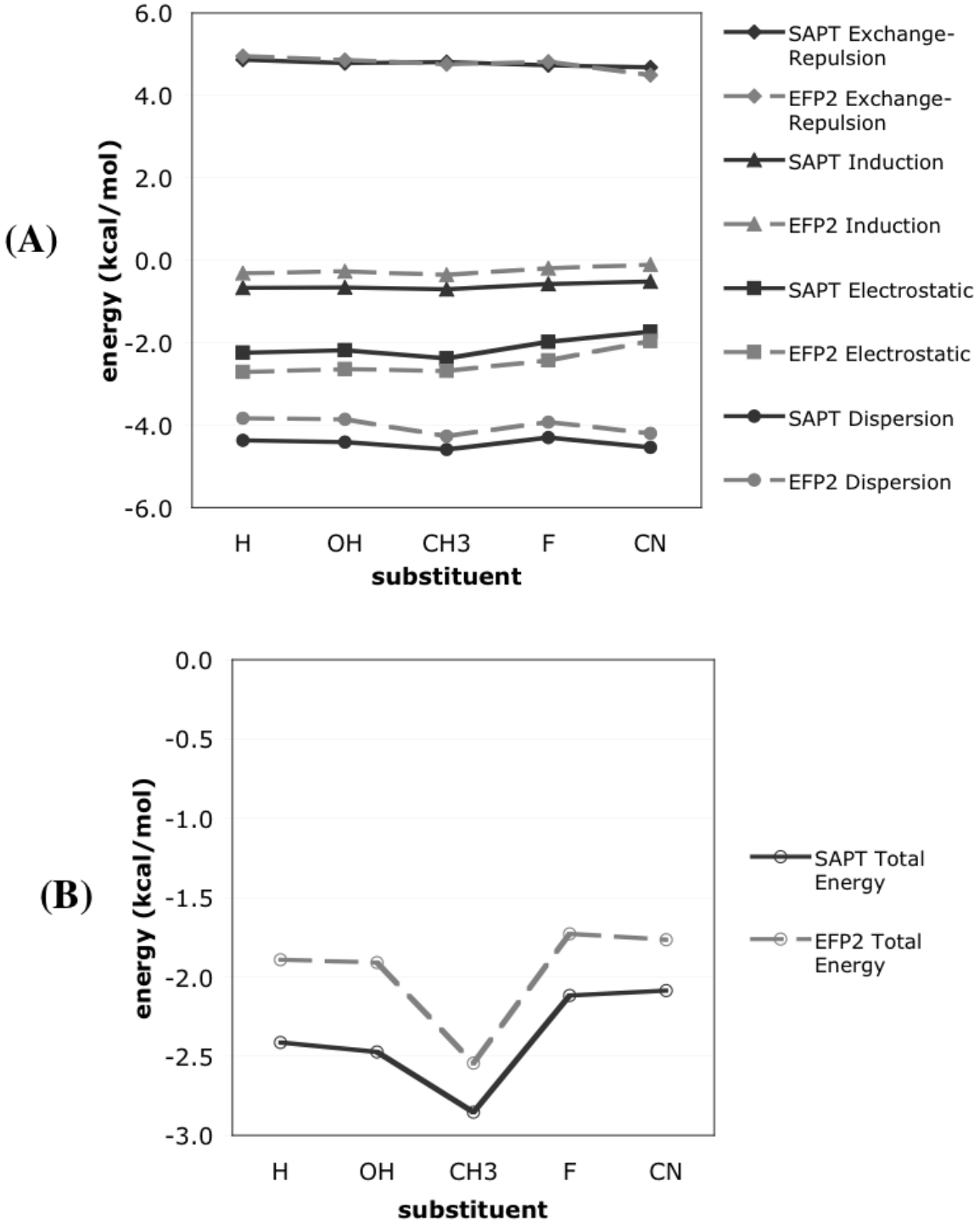


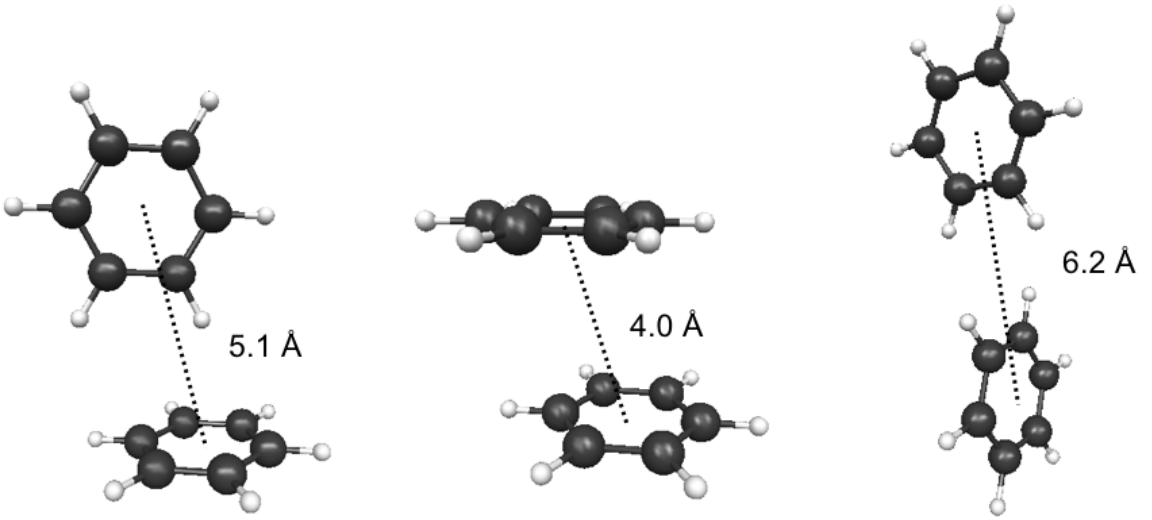
Figure 3.4. (A) Electrostatic, exchange-repulsion, polarization (induction), and dispersion components and **(B)** total interaction energies for the T-shaped(2) dimers computed with EFP2 and SAPT. SAPT data are from ref. 1.

SAPT predicts the order of increasing binding energies for the T-shaped(2) structures to be CN ≈ F < H ≈ OH < CH_3, while EFP2 predicts F < CN (Fig. 3.4A,B). However, these energy differences are very small. In general, the EFP2 exchange-repulsion energy in T-shaped(2) dimers closely resembles the SAPT exchange-repulsion. The EFP Coulomb terms are more attractive by 0.2 to 0.5 kcal mol^{-1} compared with SAPT. As with the T-shaped(1) dimers, the magnitude of the T-shaped(2) EFP2 polarization and dispersion energies are slightly lower than those for SAPT; this results in some error cancellation with the Coulomb term.

To summarize, trends due to substituent effects in each interaction energy component, as well as in the total binding energies, are well reproduced by EFP2 (Figs. 3.2-3.4). This is encouraging for future EFP2 studies of more complex species.

The potential energy surface. The conformational space of each dimer was explored using Monte Carlo/simulated annealing with the EFP2 method. The lowest energy structures for each dimer are shown in Figures 3.5-3.9. All geometries shown have positive definite Hessians.

The three minima found by EFP2 on the benzene dimer potential energy surface are parallel-displaced (Fig. 3.5B), tilted T-shaped (Fig. 3.5A), and edge-to-edge (Fig. 3.5C) structures, in agreement with recent SAPT(DFT) studies by Podeszwa, *et al*.[33] The parallel-displaced and tilted T-shaped configurations are predicted to be almost isoenergetic, with CCSD(T) favoring the latter structure by 0.1 kcal mol^{-1}.[30,33] EFP2 is in agreement with CCSD(T), predicting the tilted T-shaped structure to be the global minimum, with the parallel-displaced configuration ~0.4 kcal/mol higher in energy. The edge-to-edge configuration is ~1 kcal mol^{-1} higher in energy than the parallel-displaced one. The EFP2 parallel-displaced configuration is slightly less displaced[48] than is predicted by MP2[30], CCSD(T)[31], or SAPT(DFT)[33]. As discussed in ref. 48, the discrepancy in the R2 distance (the "displacement coordinate" defined in ref. 30) between MP2 or CCSD(T) and EFP2 is 0.4 Å. That is, the EFP2 parallel-displaced structure is less relaxed (closer to the sandwich structure) than are the MP2 or CCSD(T) isomers. This results in a slightly higher EFP2 energy for the parallel-displaced structure and consequently a slightly larger (0.4 kcal mol^{-1}) energy



ELECTROSTATIC	-1.68
REPULSION	2.29
POLARIZATION	-0.24
DISPERSION	-3.17
TOTAL	-2.80
TOTAL inc. ZPE	-2.39

(A)

ELECTROSTATIC	-0.14
REPULSION	2.93
POLARIZATION	-0.30
DISPERSION	-4.90
TOTAL	-2.41
TOTAL inc. ZPE	-2.05

(B)

ELECTROSTATIC	-1.03
REPULSION	1.54
POLARIZATION	-0.04
DISPERSION	-2.00
TOTAL	-1.53
TOTAL inc. ZPE	-1.21

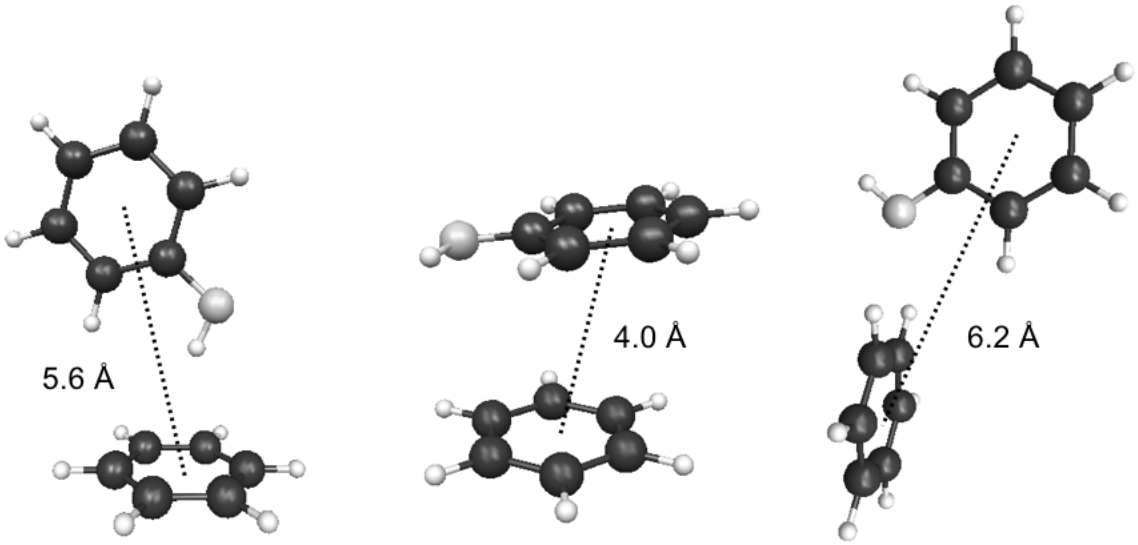
(C)

Figure 3.5. Lowest energy benzene-benzene structures found with EFP2 Monte Carlo/simulated annealing. Energies in kcal mol^{-1}.

difference between the tilted T-shaped and parallel-displaced configurations, compared to 0.1 kcal mol^{-1} for both CCSD(T) and SAPT(DFT).

For the substituted dimers, unconstrained geometry optimizations lead to minima that are similar to those of the benzene dimer; that is, configurations resembling the parallel-displaced, T-shaped, or edge-to-edge structures. However, the predicted energetic ordering of these conformations depends on the substituent. The origins of these differences are analyzed in the following discussion.

The parallel-displaced structure is common to all dimers. As in the benzene dimer, binding in the substituted parallel-displaced dimers is dominated by dispersion. As discussed above for the constrained sandwich dimers, substituents significantly increase the binding energies in the parallel-displaced dimers (by 1.3-2.8 kcal mol^{-1}). The CN-substituted dimer (Figure 3.9A) is the most strongly bound, followed by OH (Fig. 3.6B), F (Fig. 3.8A), and



ELECTROSTATIC	-3.39
REPULSION	2.76
POLARIZATION	-0.89
DISPERSION	-3.08
TOTAL	-4.60
TOTAL inc. ZPE	-4.19

(A)

ELECTROSTATIC	-1.52
REPULSION	3.83
POLARIZATION	-0.35
DISPERSION	-5.78
TOTAL	-3.82
TOTAL inc. ZPE	-3.39

(B)

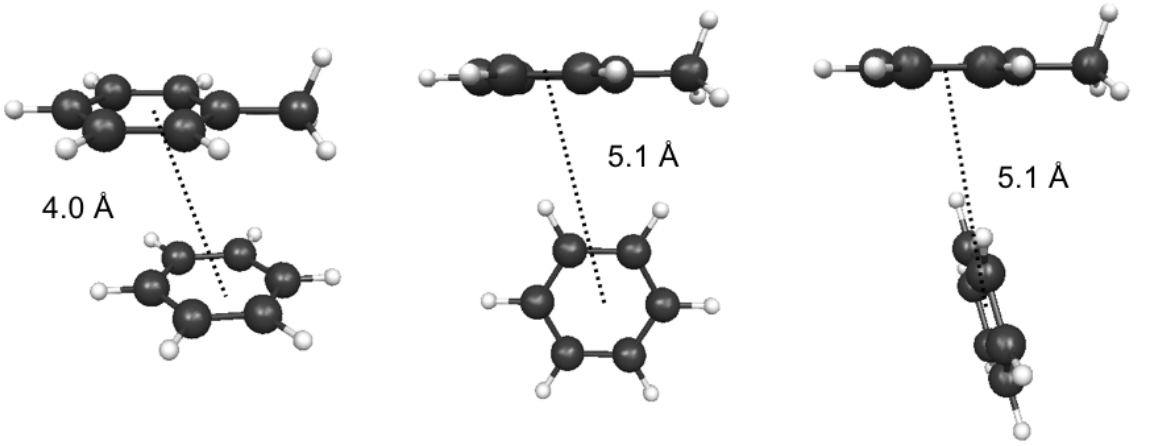
ELECTROSTATIC	-4.34
REPULSION	4.05
POLARIZATION	-0.30
DISPERSION	-2.86
TOTAL	-3.45
TOTAL inc. ZPE	-3.02

(C)

Figure 3.6. Lowest energy benzene-phenol structures, using a planar phenol monomer, found with EFP2 Monte Carlo/simulated annealing. Energies in kcal mol^{-1}.

CH_3 (Fig. 3.7A). The binding energies of the latter three structures are very similar and 1.3-1.4 kcal mol^{-1} bigger than in the parallel-displaced benzene dimer.

Binding energies in the unconstrained substituted parallel-displaced dimers are 0.6-1.1 kcal mol^{-1} stronger than those in the corresponding constrained sandwich dimers. The origin of the stronger binding is partially the electron-donating/electron-withdrawing effects of the substituents and partially the interaction between the substituent and the benzene ring in the unconstrained structures. For example, stabilization in the OH, F, and CN substituted dimers is partly due to favorable Coulomb interactions between the electronegative substituent and the positively charged hydrogens of the unsubstituted benzene. This is reflected in both the significantly larger Coulomb energies, and the slightly tilted dimer structures in which the substituent is angled closer to the benzene. In contrast, the CH_3 substituted dimer is perfectly parallel, and the additional stabilization in this species is mainly



ELECTROSTATIC	-0.78
REPULSION	3.57
POLARIZATION	-0.33
DISPERSION	-6.12
TOTAL	-3.67
TOTAL inc. ZPE	-3.25

(A)

ELECTROSTATIC	-1.74
REPULSION	2.37
POLARIZATION	-0.22
DISPERSION	-3.69
TOTAL	-3.28
TOTAL inc. ZPE	-2.87

(B)

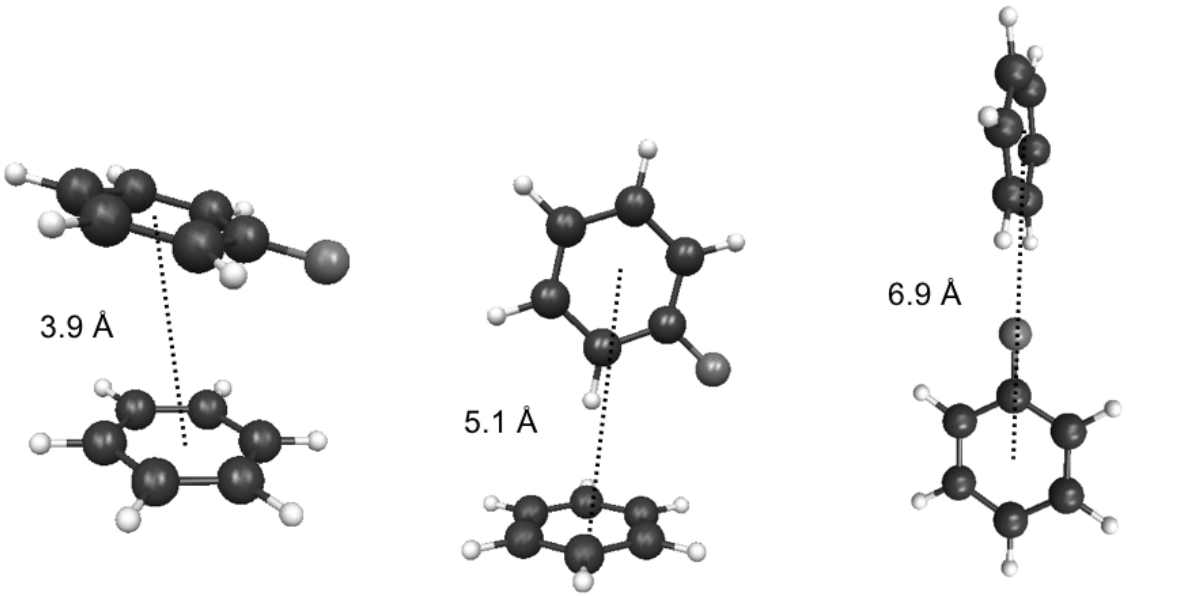
ELECTROSTATIC	-1.66
REPULSION	2.19
POLARIZATION	-0.22
DISPERSION	-3.36
TOTAL	-3.06
TOTAL inc. ZPE	-2.69

(C)

Figure 3.7. Lowest energy benzene-toluene structures found with EFP2 Monte Carlo/simulated annealing. Energies in kcal mol^{-1}.

due to the dispersion forces between the CH_3 group and the benzene ring. Consequently, the benzene-toluene dimer (Fig. 3.7A) has the smallest binding energy of the substituted parallel-displaced species.

Binding in the T-shaped dimers is governed by a balance between the dispersion and Coulomb interactions. Among the unconstrained EFP2 T-shaped structures, the CN-substituted dimer (9b) has the largest binding energy of 4.7 kcal mol^{-1}, closely followed by the OH dimer (3.6A). Two CH_3-substituted dimers (3.7B,C) and the F-substituted dimer (3.8B) have binding energies of 3.0 – 3.3 kcal mol^{-1}. As in the parallel-displaced structures, the benzene-benzonitrile and benzene-fluorobenzene dimers are stabilized through Coulomb interactions between the electronegative substituents and the unsubstituted benzene ring. The T-shaped benzene-phenol structure is different from other T-shaped dimers because its binding is dominated by the interaction between the partially positive hydrogen of the hydroxyl group and the negative π cloud of the unsubstituted benzene. The interaction in benzene-phenol is thus similar to that in the benzene-water dimer.[62] The important role of the



ELECTROSTATIC	-1.64
REPULSION	3.65
POLARIZATION	-0.24
DISPERSION	-5.55
TOTAL	-3.77
TOTAL inc. ZPE	-3.36

(A)

ELECTROSTATIC	-2.38
REPULSION	2.69
POLARIZATION	-0.31
DISPERSION	-3.28
TOTAL	-3.28
TOTAL inc. ZPE	-2.88

(B)

ELECTROSTATIC	-3.90
REPULSION	3.30
POLARIZATION	-0.30
DISPERSION	-1.87
TOTAL	-2.76
TOTAL inc. ZPE	-2.37

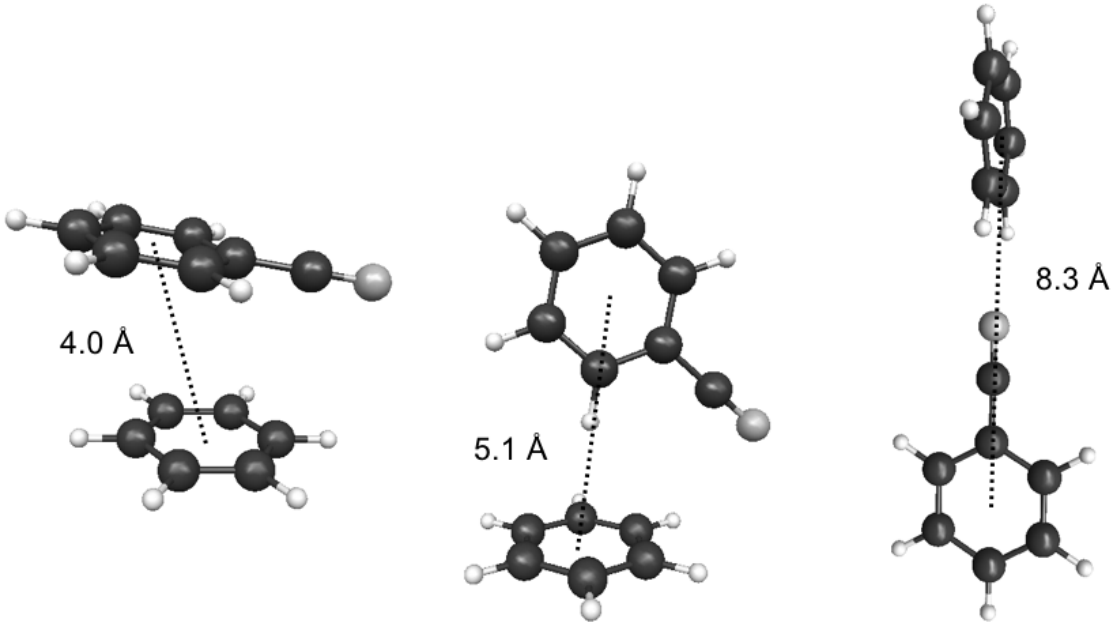
(C)

Figure 3.8. Lowest energy benzene-fluorobenzene structures found with EFP2 Monte Carlo/simulated annealing. Energies in kcal mol^{-1}.

hydroxyl group in the binding of benzene-phenol in the T-shaped isomer is also reflected in the modestly larger polarization term as compared to other dimers. Both conformations of the T-shaped-like benzene-toluene dimer are slightly stabilized by dispersion.

All substituted dimer edge-to-edge structures are also bound more strongly than is the analogous benzene dimer isomer. Additional stabilization in OH, F, and CN substituted dimers occurs due to Coulomb interactions between the positively charged hydrogens of the benzene ring and the partial negative charge on the substituent group. Indeed, in these dimers, a weak edge-to-edge interaction between the two benzene rings is augmented by a stronger edge-to-substituent interaction. As for the T-shaped dimers, the strength of the interaction in the edge-to-edge-type dimers decreases in the order CN > OH > F.

Now, consider how substituents affect the relative energies of the different conformations of the dimers. Based on the preceding analyses, it is expected that the EFP2



ELECTROSTATIC	-2.40
REPULSION	4.37
POLARIZATION	-0.34
DISPERSION	-6.83
TOTAL	-5.19
TOTAL inc. ZPE	-4.73

(A)

ELECTROSTATIC	-3.67
REPULSION	4.10
POLARIZATION	-0.67
DISPERSION	-4.48
TOTAL	-4.73
TOTAL inc. ZPE	-4.25

(B)

ELECTROSTATIC	-5.29
REPULSION	4.18
POLARIZATION	-0.93
DISPERSION	-2.32
TOTAL	-3.72
TOTAL inc. ZPE	-3.33

(C)

Figure 3.9. Lowest energy benzene-benzonitrile structures found with EFP2 Monte Carlo/simulated annealing. Energies in kcal mol^{-1}.

interactions are within 0.5 kcal mol^{-1} or less of the CCSD(T) values, so the trends are generally qualitatively and quantitatively captured by the EFP2 method. In all but benzene-benzene and phenol-benzene, the most strongly bound EFP2 structure is parallel-displaced rather than T-shaped. A rationale for this is that in the offset structures, the optimal benzene-ring orientation stays almost unchanged, but additional stabilization occurs due to favorable Coulomb and dispersion interactions of the substituent group with the unsubstituted benzene. In the substituted T-shaped dimers the interaction between the benzene rings is sterically less favorable than in the unsubstituted T-shaped benzene dimer. The exception to this trend is the benzene-phenol dimer, in which the T-shaped-like configuration is lower in energy than the parallel-displaced configuration by 0.8 kcal mol^{-1}. This is because the phenol hydroxyl group strongly interacts with the π cloud of the benzene.

Conclusions

This study explores the applicability and accuracy of the general effective fragment potential (EFP2) method, the first-principles-based model potential, to systems with aromatic π-π interactions, i.e., benzene-benzene, benzene-toluene, benzene-phenol, benzene-flourobenzene, and benzene-benzonitrile dimers. The first part of the paper compares the EFP binding energies for the constrained geometries based on the work of Sherrill and coworkers.[1] The EFP2 binding energies are in excellent agreement with the CCSD(T) results in the constrained T-shaped dimers and overestimate those in the sandwich dimers by less than 0.9 kcal mol^{-1}. For these aromatic complexes, EFP2 is more accurate than MP2. The latter, compared to CCSD(T), overestimates binding in both the T-shaped and sandwich structures by up to 1.0 and 1.8 kcal mol^{-1}, respectively. These results are excellent in view of the computational cost of each method. For example, a single-point energy calculation of the benzene dimer in the 6-311+G(3df,2p) basis with MP2 would require 142 minutes of CPU time on one IBM Power5 processor, whereas the analogous EFP2 calculation takes only 0.4 seconds.

Some of the discrepancy between EFP2 and higher levels of theory such as CCSD(T) likely occurs due to the omission of some higher-order terms in the EFP2 expansions. For instance, including induced quadrupoles in the polarization energies and higher order terms in the dispersion expansion can be of importance. These terms will be targeted in future developments of the EFP2 method.

It is very encouraging that EFP2 correctly reflects the energy changes due to different substituents. EFP2 also provides excellent agreement with symmetry adapted perturbation theory (SAPT) for the different components of the total binding energy – Coulomb, exchange-repulsion, polarization, and dispersion – demonstrating the utility of EFP2 as an accurate and computationally inexpensive tool for the analysis of binding patterns in molecular complexes and (potentially) liquids.

Finally, an independent EFP2 study was performed of the potential energy surface of each dimer by employing a Monte Carlo/simulated annealing technique. The CN, F, and CH_3

substituents stabilize and favor the parallel-displaced configurations. Binding in the benzene-phenol dimer resembles that of the water-benzene complex.

Acknowledgements

This work was supported in part by a Scientific Discovery through Advanced Computing (SciDAC) Department of Energy grant and in part by a NIRT grant from the National Science Foundation.

References

1) Sinnokrot, M.O.; Sherrill, C.D. *J. Am. Chem. Soc.* **2004**, *126*, 7690.
2) Meyer, E.A.; Castellano, R.K.; Diederich, R. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 1210.
3) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
4) Askew, B.; Ballester, P.; Buhr, C.; Jeong, K.S.; Jones, S.; Parris, K.; Williams, K.; Rebek, J, Jr. *J. Am. Chem. Soc.* **1989**, *111*, 1082.
5) Smithrud, D.B.; Diederich, F. *J. Am. Chem. Soc.* **1990**, *112*, 339.
6) Hunter, C.A. *Chem. Soc. Rev.* **1994**, *23*, 101.
7) Rebek, J., Jr. *Chem. Soc. Rev.* **1996**, *25*, 255.
8) Burley, S.K.; Petsko, G.A. *Science* **1985**, *229*, 23.
9) Hunter, C.A.; Singh, J.; Thornton, J.M. *J. Mol. Biol.* **1991**, *218*, 837.
10) Lerman, L.S. *J. Mol. Biol.* **1961**, *3*, 18.
11) Brana, M.F.; Cacho, M.; Gradillas, A.; Pascual-Teresa, B.; Ramos, A. *Curr. Pharm. Des.* **2001**, *7*, 1745.
12) Tsuzuki, S.; Uchimaru, T.; Tanabe, K. *J. Mol. Struct. (THEOCHEM)* **1994**, *307*, 107.
13) Hobza, P.; Selzle, H.L.; Schlag, E.W. *J. Phys. Chem.* **1996**, *100*, 18 790.
14) Tsuzuki, S.; Luthli, H.P. *J. Chem. Phys.* **2001**, *114*, 3949.
15) Steed, J.M.; Dixon, T.A.; Klemperer, W. *J. Chem. Phys.* **1979**, *70*, 4940.
16) Arunan, E.; Gutowsky, H.S. *J. Chem. Phys.* **1993**, *98*, 4294.
17) Law, K.S.; Schauer, M.; Bernstein, E.R. *J. Chem. Phys.* **1984**, *81*, 4871.
18) Felker, P.M.; Maxton, P.M.; Schaeffer, M.W. *Chem. Rev.* **1994**, *94*, 1787.
19) Venturo, V.A.; Felker, P.M. *J. Chem. Phys.* **1993**, *98*, 4294.
20) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104.
21) Hobza, P.; Jurecka, P. *J. Am. Chem. Soc.* **2003**, *125*, 15 608.
22) Claessens, C.G.; Stoddart, J.F.; *J. Phys. Org. Chem*. **1997**, *10*, 254.
23) Fyfe, M.C.T.; Stoddart, J.F. *Acc. Chem. Res.* **1997**, *30*, 393.
24) Dahl, T. *Acta Chem. Scand.* **1994**, *48*, 95.
25) Rappé, A.K.; Bernstein, E.R. *J. Phys. Chem. A* **2000**, *104*, 6117.

26) Kim, K.S.; Tarakeshwar, P.; Lee, J.Y. *Chem. Rev.* **2000**, *100*, 4145.
27) Dunning, T.H. *J. Phys. Chem. A* **2000**, *104*, 9062.
28) Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143.
29) Sinnokrot, M.O.; Sherrill, C.D. *J. Phys. Chem. A* **2006**, *110*, 10656.
30) Sinnokrot, M.O.; Valeev, E.F.; Sherrill, C.D. *J. Am. Chem. Soc.* **2002**, *124*, 10887.
31) Sinnokrot, M.O.; Sherrill, C.D. *J. Phys. Chem. A* **2004**, *108*, 10200.
32) Jaffe, R.L.; Smith, G.D. *J. Chem. Phys.* **1996**, *105*, 2780.
33) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345.
34) Puzder, A.; Dion, M.; Langreth, D.C. *J. Chem. Phys.* **2006**, *124*, 164105.
35) Hill, J.G.; Platts, J.A.; Werner, H.J. *Phys. Chem. Chem. Phys*. **2006**, *8*, 4072.
36) Thonhauser, T.; Puzder, A.; Langreth, D.C. *J. Chem. Phys.* **2006**, *124*, 164106.
37) Sinnokrot, M.O.; Sherrill, C.D. *J. Phys. Chem. A* **2003**, *107*, 8377.
38) DiStasio, R.A., Jr.; von Helden, G.; Steele, R.P.; Head-Gordon, M. *Chem. Phys. Lett.* **2007**, *437*, 277.
39) Tauer, T.P.; Derrick, M.E.; Sherrill, C.D. *J. Phys. Chem. A* **2005**, *109*, 191.
40) Sun, S.; Bernstein, E.R. *J. Phys. Chem.* **1996**, *100*, 13348. Also see (28)
41) Raghavachari, K.; Trucks, G.W.; Pople, J.A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.
42) Hunter, C.A.; Sanders, J.K.M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.
43) Dion, M; Rydberg, H.; Schrder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.*, **2004**, *92*, 246401.
44) Gordon, M.S.; Freitag, M.A.; Bandyopadhyay, P.; Jensen, J.H.; Kairys, V.; Stevens, W.J. *J. Phys. Chem. A* **2001**, *105*, 293.
45) Jensen, J.H.; Gordon, M.S. *Mol. Phys.* **1996**, *89*, 1313.
46) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.
47) Möller, C; Plesset, M.S. *Phys. Rev.* **1934**, *46*, 618.
48) Slipchenko, L.V.; Gordon, M.S. *J. Comp. Chem.* **2007**, *28*, 276.
49) Chen, W.; Gordon, M.S. *J. Chem. Phys.* **1996**, *105*, 11081.
50) Webb, S.P.; Gordon, M.S. *J. Phys. Chem. A* **1999**, *103*, 1265.
51) Merrill, G.N.; Gordon, M.S. *J. Phys. Chem. A* **1998**, *102*, 2650.
52) Day, P.N.; Pachter, R.; Gordon, M.S. *J. Chem. Phys.* **2000**, *112*, 2063.
53) Adamovic, I.; Gordon, M.S. *J. Phys. Chem. A*, **2005**, *109*, 1629.
54) Bandyopadhyay, P.; Gordon, M.S. *J. Chem. Phys.* **2000**, *113*, 1104.
55) Bandyopadhyay, P.; Gordon, M.S; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2002**, *116*, 5023.
56) Adamovic, I.; Li, H.; Lamm, M.H.; Gordon, M.S. *J. Phys. Chem. A*, **2006**, *110*, 519.
57) Adamovic, I.; Gordon, M.S. *J. Phys. Chem.* **2006**, *110*, 10267.
58) Slipchenko, L.V.; Gordon, M.S. (in prep.)
59) Schmidt, M.W.; Baldridge, K.K.; Boatz, J.A.; Elbert, S.T.; Gordon, M.S.; Jensen, J.H.; Koseki, S.; Matsunaga, N.; Nguyen, K.A., *et al. J Comp. Chem.* **1993**, *14*, 1347.
60) Woon, D.E.; Dunning, T.H., Jr. *J. Chem. Phys*. **1993**, *98*, 1358.
61) Day, P.N.; Pachter, R.; Gordon, M.S.; Merrill, G.N. *J. Chem. Phys*. **2000**, *112*, 2063.
62) Suzuki, S.; Green, P.G.; Bumgarner, R.E.; Dasgupta, S.; Goddard, W.A.; Blake, G.A. *Science*, **1992**, *257*, 942.

CHAPTER 4: BENZENE-PYRIDINE INTERACTIONS PREDICTED BY THE EFFECTIVE FRAGMENT POTENTIAL METHOD

Pending publication in the *Journal of Physical Chemistry A*

Q. A. Smith, L. V. Slipchenko, and M. S. Gordon

Abstract

The accurate representation of nitrogen-containing heterocycles is essential for modeling biological systems. In this study, the general effective fragment potential (EFP2) method is used to model dimers of benzene and pyridine, complexes for which high-level theoretical data – including large basis spin-component-scaled second-order perturbation theory (SCS-MP2), symmetry adapted perturbation theory (SAPT), and coupled cluster with singles, doubles, and perturbative triples (CCSD(T)) – are available. An extensive comparison of potential energy curves and components of the interaction energy is presented for sandwich, T-shaped, parallel displaced, and hydrogen bonded structures of these dimers. EFP2 and CCSD(T) potential energy curves for the sandwich, T-shaped, and hydrogen-bonded dimers have an average root mean square deviation (RMSD) of 0.49 kcal/mol; EFP2 and SCS-MP2 curves for the parallel displaced dimers have an average RMSD of 0.52 kcal/mol. Additionally, results are presented from an EFP2 Monte Carlo/simulated annealing (MC/SA) computation to sample the potential energy surface of the benzene-pyridine and pyridine dimers.

Introduction

Aromatic π-π interactions are of fundamental importance in biological systems: they appear in protein folding[1], DNA structure[2], and drug binding[3]. Benzene and substituted benzene dimers have been used as model systems to study these interactions, but biologically important molecules, including amino acids and DNA bases, often contain aromatic heterocycles in addition to substituents on the aromatic ring. Therefore, for a theoretical

method to be useful for modeling biological systems, it must be able to accurately describe the effects on π-π interactions arising from both ring substituents and from heteroatom (especially nitrogen) substitution.

The general effective fragment potential (EFP2)[4] method has previously been used to model the π-π and π-hydrogen interactions of benzene dimers[5], benzene-water complexes[6], and substituted benzene dimers[7] to a high degree of accuracy, approaching that of coupled cluster theory with singles, doubles, and perturbative triples (CCSD(T))[8]. The present work extends these benchmark EFP2 studies to examine the π-π and π-hydrogen interactions of aromatic nitrogen-containing heterocycles, specifically dimers of pyridine and benzene-pyridine. Pyridine and benzene-pyridine interactions have been examined extensively at a high level of *ab initio* theory by Hohenstein and Sherrill.[9] Also notable are previous density functional theory (DFT)[10] and combined *ab initio* and DFT investigations[11]. A favorable comparison between EFP2 and the high-level results for these chemical systems will help to establish EFP2 as a viable method for modeling more complex bio-molecules.

EFP2 is an *ab initio*-based model potential method that was designed to model intermolecular interactions. A set of parameters to account for the major noncovalent forces – Coulomb, exchange-repulsion, polarization (induction), and dispersion – for each unique fragment is derived from a single *ab initio* calculation. These parameters can be used in subsequent calculations to model interactions with other EFP fragments or with *ab initio* molecules. Unlike many other model potential methods, the parameters are not empirically fitted. EFP fragments have frozen internal geometries.

The advantage of EFP over traditional *ab inito* methods is its low computational cost. For example, a single-point energy calculation for the benzene dimer with second order perturbation theory (MP2)[10] and the 6-311++G(3df,2p) basis set[11] takes 142 minutes on a single IBM Power5 processor, while an EFP2 calculation with the same basis set takes only 0.4 seconds, after the EFP2 potential has been generated in a prior calculation.[5] The calculation to generate the EFP2 potential on a single fragment takes about as long as a MP2 calculation.

Methods

The general effective fragment potential (EFP2) method is coded in the GAMESS (General Atomic and Molecular Electronic Structure System)[13] computational chemistry software package, which was used for all calculations in this study. The EFP2 energy is comprised of Coulomb, exchange-repulsion, polarization (induction), dispersion, and charge transfer terms. The Coulomb term is computed via Stone's distributed multipolar expansion,[14] carried out through octopole moments. Both analytical and numerical[15] distributed multipolar analysis (DMA) is available; the numerical DMA was chosen for this study. Exchange-repulsion is derived as an expansion in the intermolecular overlap, truncated at the quadratic term.[16] EFP2 polarization is expressed as a sum of localized molecular orbital (LMO) polarizabilities, where the polarizable points are located at the LMO centroids (i.e., valence bonds and lone pairs of the molecule). The LMO polarizabilities are calculated from the coupled-perturbed Hartree-Fock equations.[17] Dispersion is expressed as $E_{disp} \approx \frac{C_6}{R^6} + \frac{C_8}{R^8}$, with an explicitly derived C_6/R^6 term and an estimated C_8/R^8 term. The C_6 coefficient is derived from the frequency-dependent polarizabilities integrated over the imaginary frequency range.[18] The charge transfer term, which is omitted in the present work, was derived using a perturbative analysis of the interaction between occupied orbitals on one fragment and virtual orbitals on a second fragment.[19] The charge transfer interaction is significant only if charged or highly polar species (e.g. water) are present; previous work[19] has shown that this term does not contribute significantly to the total interaction energy of most neutral molecules.

Various Coulomb damping functions are employed in the EFP2 method for modeling charge penetration.[20] In this study, charge penetration was modeled by an exponential damping function multiplying the distributed multipoles, including charge-charge, charge-dipole, charge-quadrupole, dipole-dipole, and dipole-quadrupole terms. The polarization and dispersion terms were also screened to account for close range interaction. Polarization screening takes the form of a Gaussian function and dispersion screening uses an overlap-based damping factor, both with no parameterization. These damping functions are described extensively in ref. 20.

Monomer and dimer geometries for benzene-pyridine and pyridine-pyridine dimers were obtained from ref. 9. In that work, experimental geometries were chosen for the monomers. The monomer geometry of benzene is from Gauss and Stanton[21], with a carbon-carbon bond length of 1.3915 Å and a carbon-hydrogen bond length of 1.0800 Å. The monomer geometry of pyridine is that of Innes et al.[22] EFP2 potentials for the monomers at these geometries were calculated with the 6-311++G(3df,2p) basis. In both ref. 9 and in the present study, the monomers are held rigid, having fixed internal coordinates.

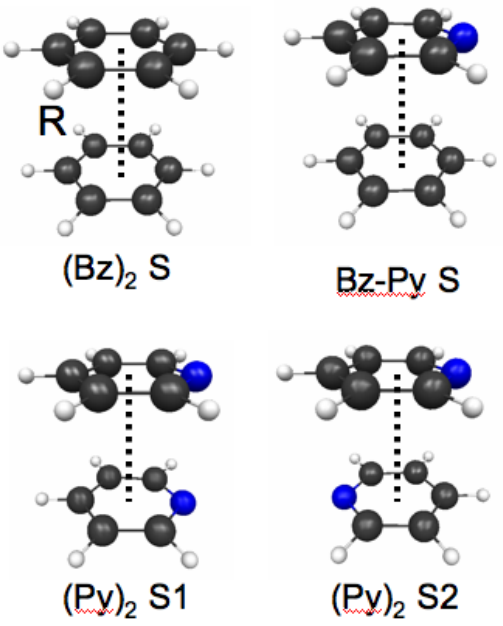


Fig. 4.1. Sandwich configurations of benzene ($(Bz)_2$), benzene-pyridine (Bz-Py), and pyridine ($(Py)_2$) dimers. R is the distance between ring centers.

Dimer geometries were also obtained from ref. 9. These geometries were chosen in order to gain a better understanding of fundamental π-π and π-hydrogen interactions and to examine how these interactions are affected by the heteroatom; therefore, the dimer geometries are not potential energy minima or experimentally obtained structures. (For example, the benzene dimer sandwich structure is a saddle point between two symmetry-equivalent parallel displaced structures.[23]) Three major classes of benzene (Bz) and pyridine (Py) dimer geometries were examined: sandwich (Fig. 4.1), T-shaped (Fig. 4.2), and parallel displaced (Fig. 4.3). For sandwich and T-shaped dimers, the distance R is a measure of the

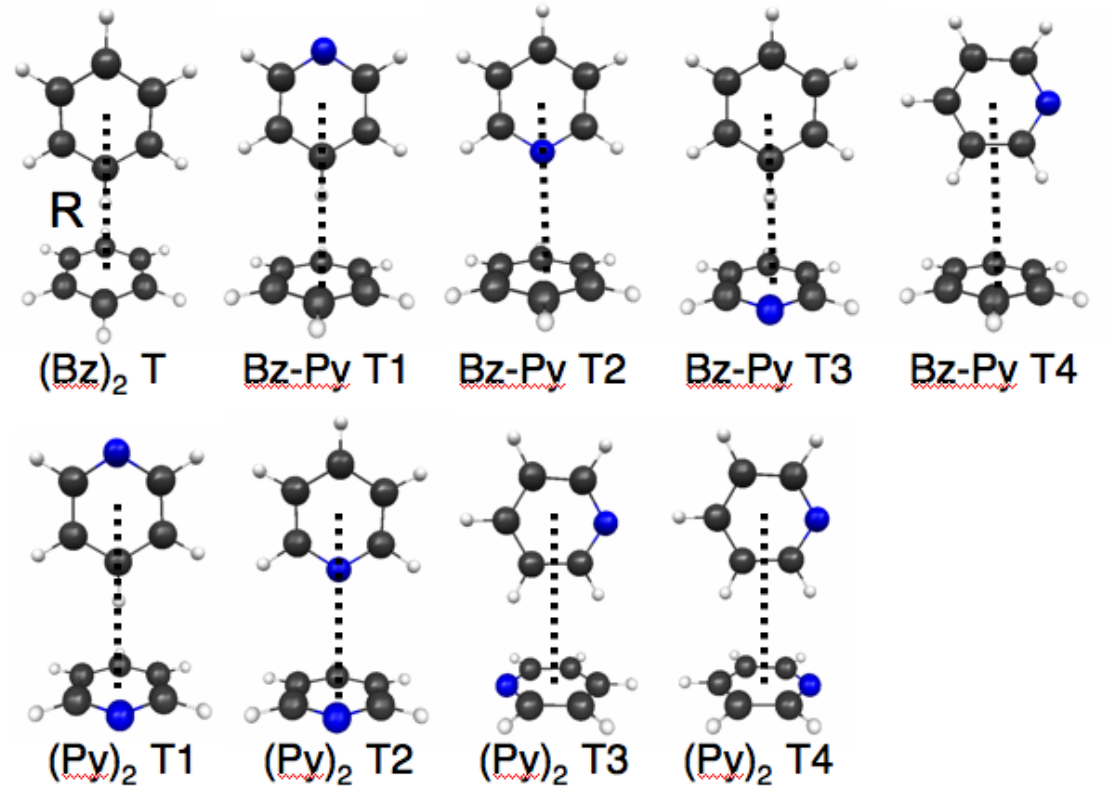


Fig. 4.2. T-shaped configurations of benzene ($(Bz)_2$), benzene-pyridine (Bz-Py), and pyridine ($(Py)_2$) dimers. R is the distance between ring centers.

vertical separation between the ring centers. Potential energy curves for these two types of dimers were obtained by varying R in increments of 0.1 Å near the potential minimum (3.2 Å $\leq$ R $\leq$ 5.0 Å for sandwich; 4.2 Å $\leq$ R $\leq$ 6.0 Å for T-shaped) and by 0.5 Å far from the minimum (5.0 Å $\leq$ R $\leq$ 7.0 Å for sandwich; 6.0 Å $\leq$ R $\leq$ 9.0 Å for T-shaped). Individual parallel displaced configurations had a vertical displacement R of 3.4 Å and a horizontal displacement H of 1.6 Å. These choices facilitate comparison with results from symmetry adapted perturbation theory (see below). To generate parallel displaced potential energy curves, the vertical displacement was fixed at 3.5 Å, while the horizontal displacement was varied in increments of 0.2 Å. Parallel displaced conformations pass through sandwich conformations at H = 0 Å. They are designated with an "a" if the horizontal displacement goes over a vertex (the heteroatom), and they are designated with a "b" if the displacement goes over an edge (the bond between pyridine carbons 2 and 3). A (+) or a (-) indicates the direction of the displacement in cases in which the displacement is not symmetric.

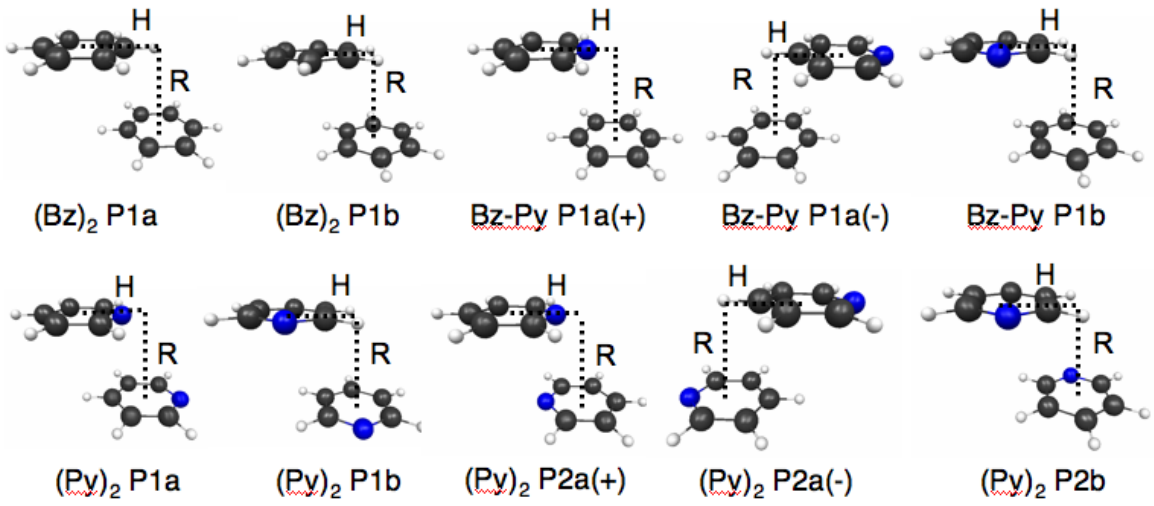


Fig. 4.3. Parallel displaced configurations of benzene ($(Bz)_2$), benzene-pyridine (Bz-Py), and pyridine ($(Py)_2$) dimers. R is the vertical separation and H is the horizontal separation between monomers.

Displacement of the upper monomer is considered to be positive if it is moved left to right relative to the perspective shown in Fig. 4.3.

Although the major focus of this study is on π-π and π-hydrogen interactions, the hydrogen bonded pyridine dimer was investigated as well (Fig. 4.4), due to the importance of hydrogen bonded interactions in biological compounds. The distance between ring centers, R, in the hydrogen bonded complex was varied in increments of 0.1 Å.

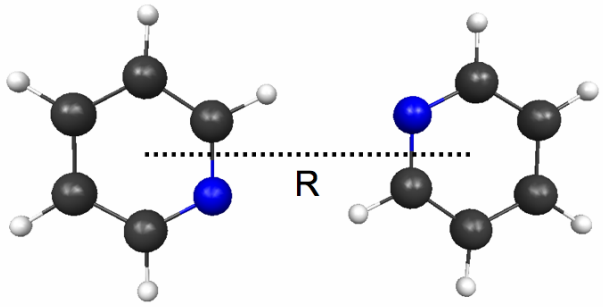


Fig. 4.4. Configuration of the hydrogen bonded pyridine dimer. R is the distance between ring centers.

EFP2 total interaction energies and potential energy curves are compared with those reported in Ref. 9. The calculations reported by Hohenstein and Sherrill[9] were performed as follows. For the sandwich, T-shaped, and hydrogen bonded dimers, large-basis CCSD(T)

single-point energies were computed by adding a coupled-cluster correction to a large-basis second-order perturbation theory (MP2) energy: $E_{CCSD(T)}^{large-basis} \approx E_{MP2}^{large-basis} + \Delta CCSD(T)$. The correction, ΔCCSD(T), is the difference between a CCSD(T) energy and an MP2 energy obtained with a smaller basis: $\Delta CCSD(T) = E_{CCSD(T)}^{small-basis} - E_{MP2}^{small-basis}$. The MP2 complete basis set (CBS) limit was estimated using the aug-cc-pVTZ and aug-cc-pVQZ basis sets[24] with the two-point extrapolation scheme of Haliker et al.[25] Energies for the parallel displaced configurations were calculated using spin-component-scaled second-order perturbation theory (SCS-MP2)[26] with the aug-cc-pVTZ basis set. The Boys-Bernardi counterpoise correction[27] was employed to account for basis set superposition error (BSSE) with all energy computations. [Note that BSSE corrections are not required for the EFP2 method.]

To analyze contributions to the total interaction energy, EFP2 Coulomb, exchange-repulsion, polarization (induction), and dispersion energy terms were determined for dimers at the geometries used by Hohenstein and Sherrill and compared with the analogous terms obtained[9] using symmetry adapted perturbation theory[28] (SAPT). Since the SAPT interaction energies obtained by Hohenstein and Sherrill are second-order or lower with respect to the intermolecular correlation operator, they are referred to as SAPT2 energies. The SAPT2 energies were computed with the aug-cc-pVDZ' basis set, which consists of the cc-pVDZ basis set with the diffuse s and p functions of aug-cc-pVDZ added to non-hydrogen atoms.

Finally, a Monte Carlo/simulated annealing (MC/SA)[29] study was performed on the EFP2 benzene-pyridine and pyridine dimers to explore their respective potential energy surfaces. Temperature ranges of both 20,000 – 100 K and 3,000 – 100 K were used. Geometry optimizations were performed every 10 steps during the MC/SA simulations.

Results and Discussion

Sandwich. A comparison of EFP2 and CCSD(T) interaction energies and intermonomer separations at the respective potential energy minima of each sandwich dimer can be found in Table 4.1. The absolute differences between EFP2 and CCSD(T) energies for the minimum-energy sandwich dimer configurations range from 0.57 kcal/mol ($(Bz)_2$ S) to 0.90 kcal/mol ($(Py)_2$ S1). For the three pyridine-containing sandwich dimers, EFP2 predicts

Table 4.1. Comparison of EFP2 and CCSD(T) potential energy minima of sandwich and T-shaped configurations of benzene, benzene-pyridine, and pyridine dimers. Intermonomer separation (R, Å) and interaction energy (ΔE, kcal/mol) of potential minima of sandwich (S) and T-shaped (T) dimers. The CCSD(T) data, taken from ref. 9, is estimated to the complete basis set (CBS) limit, except where noted by (*). The (*) dimers are est. CCSD(T)/aug-cc-pVTZ.

	CCSD(T) (Ref. 9)		EFP2	
	ΔE	R	ΔE	R
$(Bz)_2$ S	-1.76	3.9	-2.33	3.9
Bz-Py S	-2.22	3.8	-3.01	3.9
$(Py)_2$ S1	-1.61	3.8	-2.38	3.9
$(Py)_2$ S2	-2.95	3.7	-3.85	3.8
$(Bz)_2$ T	-2.73	5.0	-3.01	5.1
Bz-Py T1	-3.18	4.9	-3.34	5.2
Bz-Py T2 *	-0.64	4.7	0.08	4.8
Bz-Py T3	-2.20	5.0	-2.53	5.0
Bz-Py T4	-2.74	5.0	-2.86	5.1
$(Py)_2$ T1	-2.46	5.0	-2.69	5.0
$(Py)_2$ T2 *	-1.23	4.6	-0.45	4.8
$(Py)_2$ T3	-2.95	4.9	-2.92	5.1
$(Py)_2$ T4 *	-2.15	5.0	-2.14	5.1

an optimum intermonomer separation 0.1 Å wider than that found with CCSD(T). The EFP2 optimum separation for $(Bz)_2$ S (3.9 Å) is identical to that found with CCSD(T). Compared to the high-level CCSD(T) calculations[9], the EFP2 potential energy curves (PECs) for the sandwich dimers are deeper (dimers are more strongly bound), by up to ~0.9 kcal/mol. EFP2 and CCSD(T) predict a similar trend in the PEC ordering of the sandwich dimers (Fig. 4.5). A slight difference occurs where the EFP2 PEC for $(Py)_2$ S1 drops below that of $(Bz)_2$ S; CCSD(T) calculations predict that the benzene sandwich dimer is always more strongly bound than $(Py)_2$ S1. The EFP2 potential energy minimum for $(Py)_2$ S1 is found to be -2.38 kcal/mol at a vertical separation of 3.9 Å, while that of the benzene sandwich dimer is -2.33 kcal/mol at 3.9 Å. With CCSD(T), $(Py)_2$ S1 has a potential minimum of -1.61 kcal/mol at 3.8 Å and $(Bz)_2$ S has a minimum of -1.76 kcal/mol at 3.9 Å.[9]

The root mean square deviation (RMSD) and maximum absolute (unsigned) difference in energy (MAX) between the EFP2 PEC and the corresponding CCSD(T) PEC for each sandwich structure are shown in Table 4.2. For each set of PECs, these values were computed on a subset of the PEC data beginning from the first negative (attractive) binding energy found with EFP2 and ending with the final computed binding energy at 7.0 Å. This

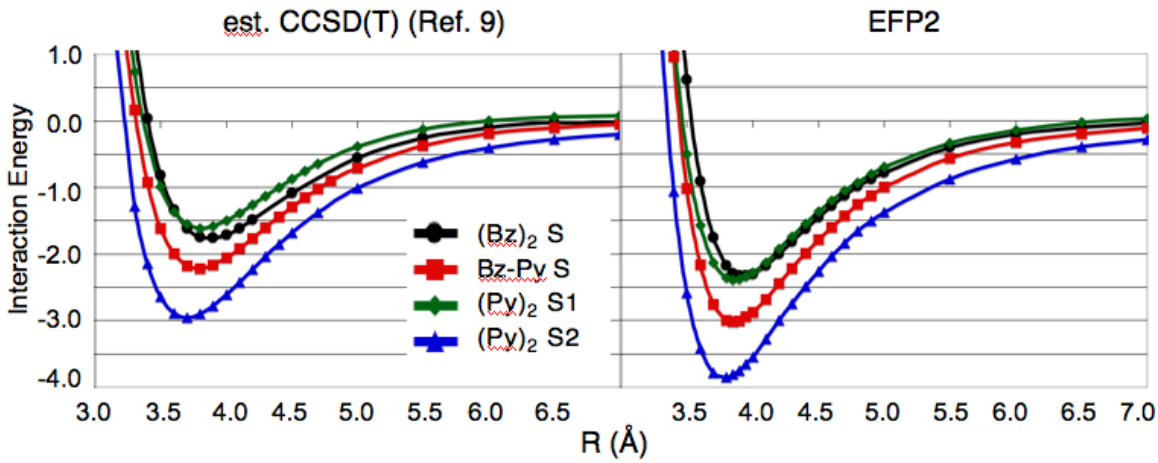
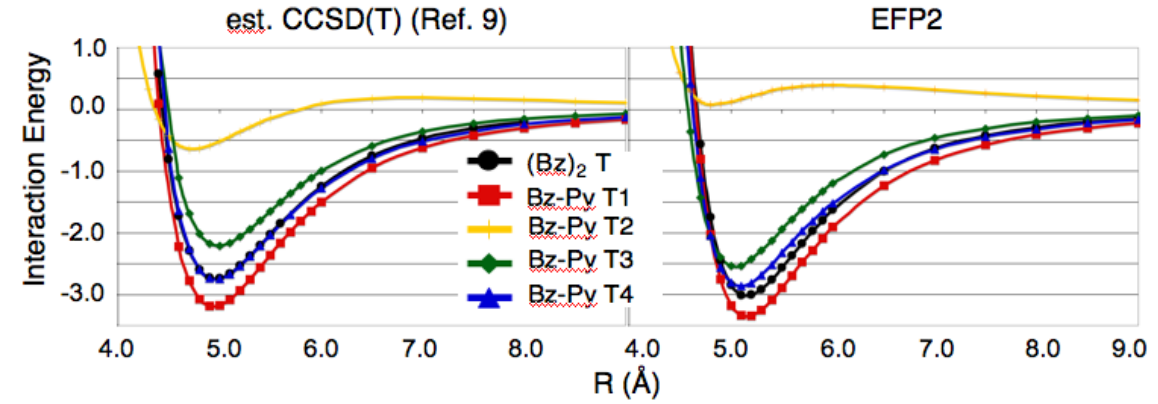


Fig. 4.5. Est. CCSD(T)/CBS[9] (left) and EFP2 (right) potential energy curves for the sandwich configurations. Interaction energy is in kcal/mol.

(A)



(B)

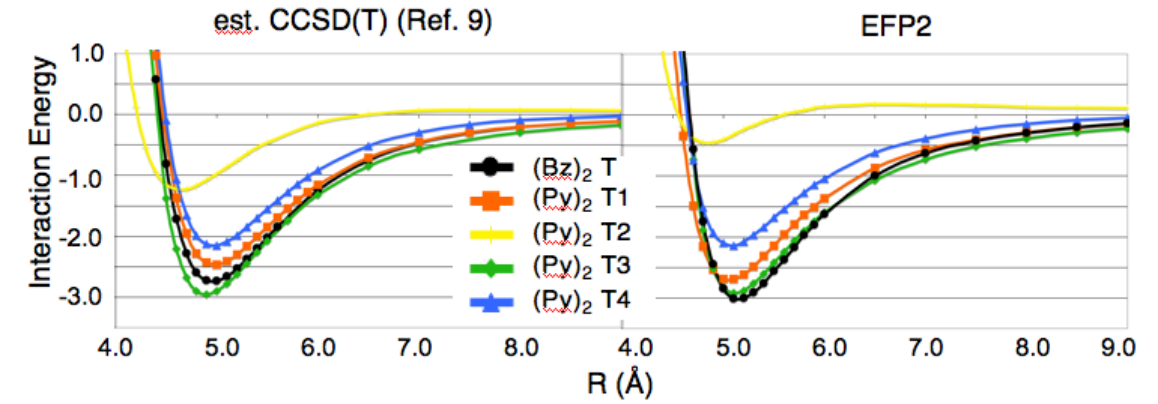


Fig. 4.6. Est. CCSD(T)/CBS[9] (left) and EFP2 (right) potential energy curves for the benzene-pyridine **(A)** and pyridine **(B)** T-shaped dimers. Interaction energy is in kcal/mol.

Table 4.2. Comparison of EFP2 and high-level *ab initio* potential energy curves. Root mean square deviation (RMSD) and maximum absolute (unsigned) energy difference (MAX) of EFP2 potential energy curves (PECs) compared to the high-level *ab initio* PECs of ref. 9. For sandwich (S), T-shaped (T), and hydrogen bonded dimers, the reference PECs were computed with CCSD(T) estimated to the complete basis set (CBS) limit, except where noted by (*). The (*) dimers are est. CCSD(T)/aug-cc-pVTZ. For parallel displaced dimers, the reference PECs were computed with SCS-MP2/aug-cc-pVTZ. Values are in kcal/mol.

	RMSD	MAX
(Bz)2 S	0.39	0.59
Bz-Py S	0.53	0.84
(Py)2 S1	0.52	0.80
(Py)2 S2	0.66	1.08
(Bz)2 T	0.61	1.72
Bz-Py T1	0.64	1.97
Bz-Py T2	0.65	1.34
Bz-Py T3	0.31	0.47
Bz-Py T4	0.40	1.16
(Py)2 T1	0.34	1.03
(Py)2 T2	0.54	1.06
(Py)2 T3	0.64	1.94
(Py)2 T4	0.27	0.92
(Bz)2 P1a	0.70	1.31
(Bz)2 P1b	0.75	1.31
Bz-Py P1a	0.55	0.95
Bz-Py P1b	0.48	0.96
(Py)2 P1a	0.40	0.68
(Py)2 P1b	0.44	0.94
(Py)2 P2a	0.47	0.87
(Py)2 P2b	0.39	0.83
(Py)2 H-bonded *	0.33	0.50

subset was chosen to give the most meaningful comparison between EFP2 and CCSD(T), since the EFP2 PECs are significantly more repulsive than the CCSD(T) PECs at short intermonomer separations (closer than ~3.6 Å for the sandwich dimers). The $(Bz)_2$ S PEC is in the closest agreement with CCSD(T), having the lowest RMSD (0.39 kcal/mol) and lowest MAX (0.59 kcal/mol). The greatest RMSD (0.66 kcal/mol) and MAX (1.08 kcal/mol) correspond to the $(Py)_2$ S2 PEC.

The EFP2 and SAPT2 contributions to the sandwich interaction energies at an intermonomer separation of 3.8 Å are summarized in Table 4.3. Since this is the distance at which the SAPT analyses are available[9], it was chosen for consistency. Consequently, the EFP2 total interaction energies that are listed in Tables 4.1 and 4.3 are different. Those in

Table 4.3. EFP2 and SAPT2 contributions to the interaction energy of the sandwich configurations of benzene, benzene-pyridine, and pyridine dimers. All sandwich dimers held at vertical separation R = 3.8 Å. Energies in kcal/mol. SAPT2 data, taken from ref. 9, calculated with the aug-cc-pVDZ' basis set.

	Coulomb	Pol	Exch-Rep	Dispersion	Total
$(Bz)_2$ S SAPT2	-0.48	-0.28	4.52	-5.68	-1.92
EFP2	-0.26	-0.36	4.38	-5.93	-2.17
Bz-Py S SAPT2	-0.80	-0.26	4.00	-5.34	-2.40
EFP2	-0.86	-0.27	3.66	-5.53	-3.00
$(Py)_2$ S1 SAPT2	-0.05	-0.21	3.57	-5.00	-1.69
EFP2	-0.25	-0.18	3.26	-5.19	-2.36
$(Py)_2$ S2 SAPT2	-1.29	-0.25	3.49	-5.00	-3.05
EFP2	-1.57	-0.20	3.08	-5.16	-3.85

Table 4.1 are more quantitatively meaningful, since they correspond to potential energy minima on the respective potential energy surfaces. The predominant attractive term in sandwich dimer interactions is the dispersion energy, due to direct overlap of the π electron clouds. However, this overlap also produces a large exchange-repulsion energy. Compared to benzene, pyridine has a more contracted, less polarizable π cloud. This causes dispersion interactions among pyridine-containing sandwich dimers to be slightly less favorable to binding, but it also decreases the magnitude of the exchange-repulsion term. As shown in Table 4.3, the $(Py)_2$ sandwich configurations have dispersion energies that are approximately 0.3 kcal/mol higher (less stabilized) than those of the Bz-Py sandwich dimer, which in turn has a dispersion energy that is approximately 0.3 kcal/mol (with SAPT2) or 0.4 kcal/mol (with EFP2) higher than the $(Bz)_2$ sandwich dimer. However, the exchange-repulsion energy is reduced accordingly. Overall, this decrease in the exchange-repulsion interaction energy makes pyridine-containing sandwich dimers more likely to be more strongly bound than the benzene sandwich dimer, despite the concomitant decrease in favorable dispersion interactions.

The exception to the above analysis is $(Py)_2$ S1, which is the least strongly bound sandwich dimer except at short intermonomer distances ($R \leq 3.6$ Å with estimated CCSD(T)/CBS; $R \leq 3.9$ Å with EFP2), where it becomes slightly more strongly bound than the benzene sandwich dimer. Even though the dispersion and exchange-repulsion terms associated with $(Py)_2$ S1 follow the trends described above, the Coulomb term is only

marginally favorable to binding, due to parallel dipole-dipole interactions. The Coulomb term in $(Py)_2$ S1 contributes only -0.05 kcal/mol (SAPT2) or -0.25 kcal/mol (EFP2) to the total interaction energy at an intermonomer separation R = 3.8 Å. This is less than or equal to the Coulomb term in the benzene dimer, which is -0.48 kcal/mol (SAPT2) or -0.26 kcal/mol (EFP2) at R = 3.8 Å. In contrast, $(Py)_2$ S2, which has antiparallel dipole-dipole interactions, exhibits a significantly stronger Coulomb interaction of -1.29 kcal/mol (SAPT2) or -1.57 kcal/mol (EFP2) at the same separation. The Coulomb term in Bz-Py S falls in between that of $(Bz)_2$ S and $(Py)_2$ S1 at both the SAPT2 and EFP2 levels of theory (Table 4.3).

T-shaped. The interaction energies for the T-shaped dimers are also listed in Table 4.1. The absolute differences between EFP2 and CCSD(T) energies (Table 4.1) for the T-shaped dimers range from 0.01 kcal/mol ($(Py)_2$ T4) to 0.78 kcal/mol ($(Py)_2$ T2). As with the sandwich structures, EFP2 tends to favor a slightly larger (0.1-0.2 Å) intermonomer separation for the T-shaped dimers compared to CCSD(T). For one dimer (Bz-Py T1), EFP2 finds an optimum separation 0.3 Å wider than CCSD(T); for two dimers (Bz-Py T3 and $(Py)_2$ T1), EFP2 and CCSD(T) optimum intermonomer separations agree exactly.

As described in ref. 9, binding in the T-shaped dimers is stabilized by the Coulomb energy due to the role of pyridine as a "π-hydrogen bond" donor. By pulling electron density away from the hydrogen atom that is para to the heteroatom, the heteroatom in pyridine gives this para hydrogen a greater positive charge compared to its value in benzene. The Coulomb contribution to the binding energy in the pyridine-containing T-shaped dimers comes largely from the interaction of the positive para hydrogen atom on one monomer with the negative π cloud of the other monomer ring. The same change in electron density that makes the pyridine para hydrogen more positive also makes the pyridine π cloud less diffuse, decreasing its ability to act as a "π-hydrogen bond" acceptor. This is reflected in the lower (in magnitude) binding energies of $(Py)_2$ T1 and T4 complexes relative to the corresponding Bz-Py T-shaped complexes.

The contributions to the T-shaped interaction energies follow a similar pattern to those for the sandwich isomers, as illustrated in Table 4.4. Note that, as discussed above in regard to Table 4.3, the distances used for the analysis in Table 4.4 are different from those in Table 1, and therefore, the total interaction energies differ as well. While the dispersion

Table 4.4. EFP2 and SAPT2 contributions to the interaction energy of the T-shaped configurations of benzene, benzene-pyridine, and pyridine dimers. Monomer center to monomer center distance *R* given in Å. Energies given in kcal/mol. SAPT2 data, taken from ref. 9, calculated with the aug-cc-pVDZ' basis set.

	R	Coulomb	Pol	Exch-Rep	Dispersion	Total
$(Bz)_2$ T SAPT2	5.0	-1.75	-0.52	3.52	-3.73	-2.48
EFP2		-2.94	-0.35	5.04	-4.59	-2.84
Bz-Py T1 SAPT2	5.0	-2.12	-0.64	3.54	-3.70	-2.92
EFP2		-3.32	-0.45	4.94	-4.34	-3.17
Bz-Py T2 SAPT2	4.7	0.33	-0.62	3.38	-3.75	-0.66
EFP2		1.97	-0.59	3.10	-4.35	0.13
Bz-Py T3 SAPT2	5.0	-1.21	-0.40	3.27	-3.51	-1.85
EFP2		-1.77	-0.15	3.07	-3.68	-2.53
Bz-Py T4 SAPT2	5.0	-1.80	-0.49	3.21	-3.36	-2.44
EFP2		-2.06	-0.29	2.91	-3.36	-2.80
$(Py)_2$ T1 SAPT2	5.0	-1.39	-0.50	3.29	-3.48	-2.08
EFP2		-2.00	-0.23	3.04	-3.50	-2.69
$(Py)_2$ T2 SAPT2	4.7	-0.39	-0.54	3.17	-3.53	-1.29
EFP2		1.14	-0.56	3.74	-4.70	-0.38
$(Py)_2$ T3 SAPT2	5.0	-1.78	-0.38	2.67	-3.11	-2.60
EFP2		-2.11	-0.23	3.12	-3.61	-2.83
$(Py)_2$ T4 SAPT2	5.0	-1.14	-0.38	2.78	-3.12	-1.86
EFP2		-1.10	-0.19	2.34	-3.15	-2.10

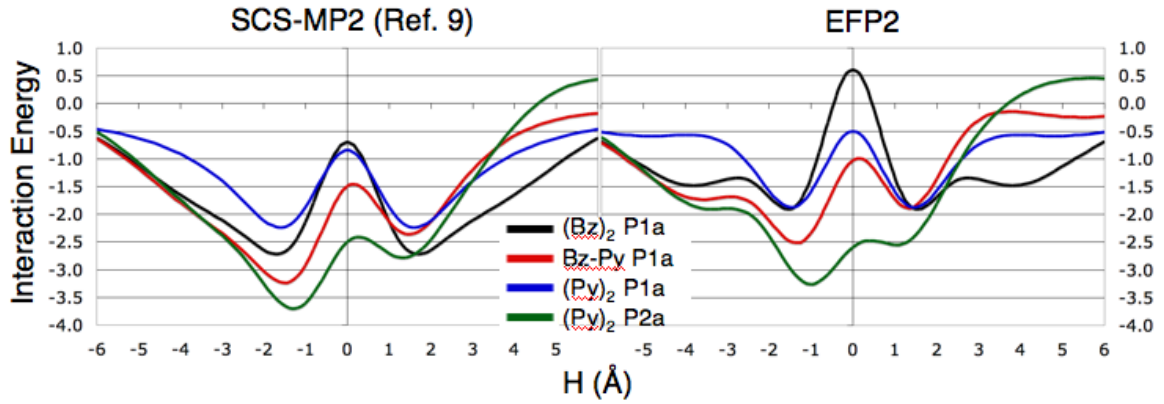
energy is the largest attractive energy term for the T-shaped dimers (Table 4.4), the exchange-repulsion energy generally has a similar magnitude. The Coulomb term consequently makes significant contributions to the overall binding energy. When the Coulomb term is least attractive or is even repulsive (T2 complexes, in which the partially negative heteroatom interacts most directly with the π cloud of the other monomer ring), the complex may be very weakly bound, or not bound at all. EFP2 predicts a more repulsive Coulomb interaction in the T2 complexes compared to SAPT2, and therefore a more weakly bound dimer. Because the EFP2 Bz-Py T2 Coulomb interaction is more than 1 kcal/mol larger than that predicted by SAPT2[9], EFP predicts this species to be unbound, whereas it is slightly (0.66 kcal/mol) bound according to SAPT2. EFP2 predicts $(Py)_2$ T2 to be more weakly bound than does SAPT2 for similar reasons. As shown in Table 1, the estimated CCSD(T) interaction energies are similar to those predicted by SAPT2. For the other T complexes, EFP2 and SAPT2 are in good agreement, within ~0.2 to 0.6 kcal/mol (Table 4.4), with the EFP2 dimers more strongly bound. Each of the contributing interactions is also in

good agreement. In these dimers, EFP2 tends to overestimate the Coulomb attraction to some degree compared to SAPT2 and to slightly (~0.2 kcal/mol) underestimate the magnitude of the polarization.

A comparison of the CCSD(T)[9] and EFP2 PECs for the T-shaped dimers is shown in Fig. 4.6. The EFP2 curves tend to be somewhat deeper (more strongly bound) and slightly right-shifted (larger intermonomer separation) compared to the CCSD(T) curves. This is consistent with the results of previous EFP2 studies on benzene[5] and substituted benzene[7] dimers. The ordering of the EFP2 and CCSD(T) curves is very similar, although EFP2 overbinds the T-shaped benzene dimer somewhat more than it does the other T-shaped dimers, resulting in some minor differences. The CCSD(T) Bz-Py T4 curve, for example, nearly coincides with the CCSD(T) $(Bz)_2$ T curve, while the EFP2 $(Bz)_2$ T curve is slightly deeper than that of Bz-Py T4. Similarly, whereas CCSD(T) shows the $(Bz)_2$ T curve always above $(Py)_2$ T4, the two nearly coincide when EFP2 is used.

RMSD values and maximum energy differences (MAX) between EFP2 and CCSD(T) T-shaped dimer PECs appear in Table 4.2. As with the sandwich dimers, these statistics were calculated on a subset of data beginning from the first negative (attractive) EFP2 binding energy for each dimer and ending with the final computed binding energy (at 8.0 Å for T-shaped dimers). In the case of Bz-Py T2, for which EFP2 finds no net binding, statistics were calculated beginning with the first attractive CCSD(T) binding energy. $(Py)_2$ T4 shows the best agreement between EFP2 and CCSD(T) curves, with a RMSD of 0.27 kcal/mol and a MAX of 0.92 kcal/mol. $(Bz)_2$ T, Bz-Py T1, Bz-Py T2, and $(Py)_2$ T3 show the greatest discrepancy between their EFP2 and CCSD(T) PECs, with RMSDs in the range of 0.61-0.65 kcal/mol. The greatest maximum difference between EFP2 and CCSD(T) energies (1.97 kcal/mol) is found in the Bz-Py T1 dimer. However, much of the difference between EFP2 and CCSD(T) energies results from EFP2 under-binding at short (equilibrium distance minus 0.4 Å, or less) intermonomer separations. If RMSDs and maximum energy differences are calculated beginning from 0.3 Å closer than the EFP2 equilibrium distance, for example, the Bz-Py T1 dimer's RMSD decreases from 0.64 to 0.39 kcal/mol and the MAX decreases from 1.97 to 0.53 kcal/mol.

(A)



(B)

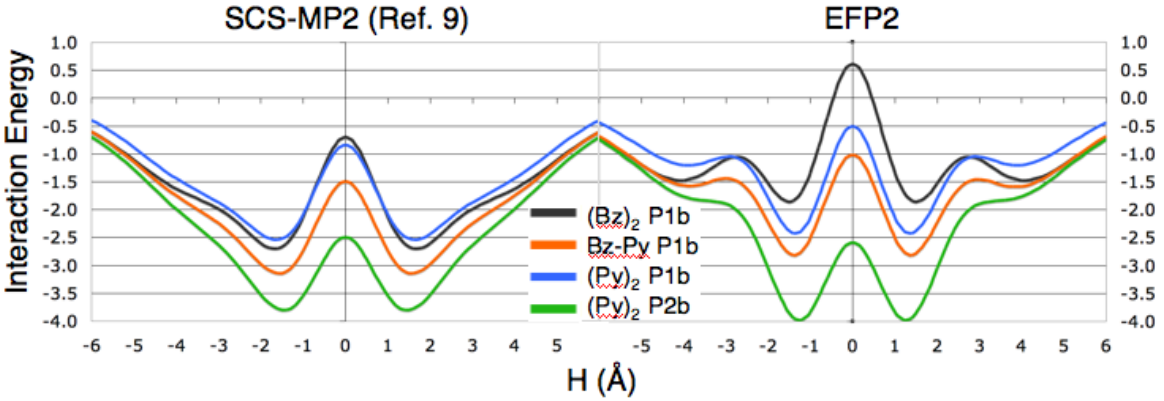


Fig. 4.7. SCS-MP2/aug-cc-pVTZ[9] (left) and EFP2 (right) potential energy curves for the parallel displaced dimers with a vertical separation R = 3.5 Å. **(A)** "Over vertex" displaced dimers, in which the displacement is over the heteroatom. **(B)** "Edgewise" displaced dimers, in which the displacement is over the C-C bond of carbons 2 and 3. Interaction energy is in kcal/mol.

Parallel displaced. Potential energy curves for the parallel displaced dimers are shown in Fig. 4.7, where EFP2 interaction energies are compared with SCS-MP2/aug-cc-pVTZ interaction energies from ref. 9. These curves correspond to parallel displaced dimers with a vertical displacement held fixed at R = 3.5 Å while the horizontal displacement, H, was varied in increments of 0.2 Å. Interaction energies and horizontal displacements of the most favorable of these structures appear in Table 4.5. The absolute energy differences between EFP2 and SCS-MP2 for these parallel displaced structures (Table 4.5) range from

0.14 kcal/mol ($(Py)_2$ P2b) to 0.86 kcal/mol ($(Bz)_2$ P1b), with an average difference of 0.50 kcal/mol.

RMSD values and maximum absolute energy differences between EFP2 and SCS-MP2 PECs for the parallel displaced dimers can be found in Table 4.2. Of the EFP2 parallel displaced PECs, those for the benzene dimers ($(Bz)_2$ P1a and $(Bz)_2$ P1b) differ the most from the corresponding SCS-MP2 PECs, having RMSDs of 0.70 and 0.75 kcal/mol respectively. The maximum absolute energy difference is 1.31 kcal/mol for both $(Bz)_2$ PECs. RMSD values for the pyridine-containing parallel displaced dimers are lower, ranging from 0.39 kcal/mol ($(Py)_2$ P2b) to 0.55 kcal/mol (Bz-Py P1a). The maximum energy differences for pyridine-containing dimers are all under 1 kcal/mol, with $(Py)_2$ P1a having the lowest maximum difference at 0.68 kcal/mol.

Table 4.5. EFP2 and SCS-MP2 interaction energies of parallel displaced configurations of benzene, benzene-pyridine, and pyridine dimers. Vertical displacement R = 3.5 Å. Horizontal displacement H given in Å. Energies given in kcal/mol. SCS-MP2/aug-cc-pVTZ data from ref. 9.

	H	SCS-MP2	EFP2
$(Bz)_2$ P1a	1.6	-2.71	-1.90
$(Bz)_2$ P1b	1.6	-2.70	-1.84
Bz-Py P1a(+)	1.4	-2.36	-1.88
Bz-Py P1a(-)	1.6	-3.23	-2.43
Bz-Py P1b	1.6	-3.14	-2.71
$(Py)_2$ P1a	1.6	-2.24	-1.85
$(Py)_2$ P1b	1.6	-2.54	-2.31
$(Py)_2$ P2a(+)	1.4	-2.78	-2.44
$(Py)_2$ P2a(-)	1.2	-3.70	-3.21
$(Py)_2$ P2b	1.4	-3.80	-3.94

Changes in the curvature of EFP2 potential energy curves for the parallel displaced structures (Fig. 4.7) can be seen at horizontal separations in the range of 2-4 Å (and -2 to -4 Å). In some cases, these curvature changes give rise to small dips in the potential energy surface that do not correspond to features on the SCS-MP2 potential energy curves (Fig. 4.7). However, the dips are well above the minimum energies found along these curves, and the overall correspondence between the EFP2 and SCS-MP2 curves is strong (Table 4.2).

The most favorable structure, predicted by both EFP2 and SCS-MP2[9], is $(Py)_2$ P2b, due to its antiparallel dipoles. For all other parallel displaced structures around their energy

minima, EFP2 predicts a somewhat smaller binding energy than does SCS-MP2, but EFP2 slightly overbinds the $(Py)_2$ P2b structure by ~0.14 kcal/mol. The next most favorable parallel displaced configuration, $(Py)_2$ P2a(-), is about 0.2 (0.7) kcal/mol higher in energy according to SCS-MP2 (EFP2). The Bz-Py P1a(-) and Bz-Py P1b complexes are next lowest in energy. Both SCS-MP2 and EFP2 predict these two dimers to be very close to each other in energy. While SCS-MP2 predicts that $(Py)_2$ P1a and $(Py)_2$ P1b have a smaller binding energy than $(Bz)_2$ P1a and $(Bz)_2$ P1b, respectively (due to parallel dipoles), EFP2 under-binds the $(Bz)_2$ dimers and reverses this energy ordering.

EFP2 tends to favor a slightly larger (0.05 - 0.25 Å) intermonomer separation in sandwich-type structures: it underestimates the Coulomb attraction and overestimates the exchange-repulsion at shorter distances.[5,7] (See also the optimal intermonomer distances listed in Table 4.1.) It was demonstrated previously[5] that EFP2 will produce parallel displaced benzene dimer interaction energies that are closer to the CCSD(T) values when the EFP2 structures are allowed to relax to their slightly more separated equilibrium values. For example, a previous comparison[5] with CCSD(T) potential energy curves for the parallel displaced benzene dimer[30] shows that the $(Bz)_2$ equilibrium geometry is R = 3.6 Å, H = 1.6 Å with CCSD(T) but R = 3.8 Å, H = 1.2 Å with EFP2. As a result, the EFP2 curve at R = 3.4 Å lies 2.5 - 4.0 kcal/mol higher in energy than the CCSD(T) curve. However, the EFP2 curve at R = 3.8 Å follows the corresponding CCSD(T) curve very closely.[5] The same may be true for heteroatom-containing dimers as well.

EFP2 and SAPT2 energy decompositions for the parallel displaced configurations with vertical displacement R = 3.4 Å and horizontal displacement H = 1.6 Å are given in Table 4.6. In light of the above discussion, this intermonomer separation may be well below the optimal separation found with EFP2; consequently, the agreement here with SAPT2 is not as favorable, but this is the only intermonomer separation for which the SAPT analysis is available. Compared to SAPT2, EFP2 tends to underestimate the Coulomb attraction in these dimers by about 0.5 kcal/mol. The closest agreement is in $(Py)_2$ P2b, where the EFP2 underestimation is 0.28 kcal/mol, and the greatest difference is in $(Py)_2$ P2a(+), with an EFP2 underestimation of 1.22 kcal/mol. The EFP2 polarization attraction is about half that of the SAPT2 polarization. EFP2 overestimates the exchange-repulsion of the $(Bz)_2$ dimers by more

than 1 kcal/mol. These dimers are the least strongly bound and have an EFP2 total interaction energy differing from the SAPT2 total interaction energy by about 2 kcal/mol (less strongly bound with EFP2). The EFP2 and SAPT2 comparisons for Bz-Py P1a(-) and $(Py)_2$ P2a(-) are similar, with an exchange-repulsion 0.96 kcal/mol and 0.87 kcal/mol (respectively) higher with EFP2 and a total interaction energy 1.87 kcal/mol and 1.72 kcal/mol (respectively) less strongly bound with EFP2. Otherwise, the EFP2 total interaction energies differ from those of SAPT2 by less than 1.4 kcal/mol. EFP2 dispersion energies are in excellent agreement with SAPT2, being consistently about 0.2 kcal/mol more attractive.

Table 4.6. EFP2 and SAPT2 contributions to the interaction energy of the parallel displaced configurations of benzene, benzene-pyridine, and pyridine dimers. Vertical displacement R = 3.4 Å, horizontal displacement H = 1.6 Å. Energies given in kcal/mol. SAPT2 data from ref. 9.

	Coulomb	Pol	Exch-Rep	Dispersion	Total
(Bz)2 P1a SAPT2	-2.77	-0.88	8.58	-7.88	-2.95
EFP2	-2.09	-0.47	9.70	-8.12	-0.98
(Bz)2 P1b SAPT2	-2.81	-0.91	8.68	-7.88	-2.92
EFP2	-2.10	-0.47	9.79	-8.12	-0.90
Bz-Py P1a(+) SAPT2	-1.91	-0.73	7.24	-7.16	-2.55
EFP2	-1.01	-0.36	7.60	-7.39	-1.16
Bz-Py P1a(-) SAPT2	-3.24	-0.85	8.25	-7.59	-3.43
EFP2	-2.60	-0.38	9.21	-7.79	-1.56
Bz-Py P1b SAPT2	-2.95	-0.81	7.77	-7.38	-3.37
EFP2	-2.49	-0.36	8.45	-7.62	-2.02
(Py)2 P1a SAPT2	-1.74	-0.68	6.97	-6.90	-2.35
EFP2	-1.05	-0.27	7.15	-7.07	-1.24
(Py)2 P1b SAPT2	-2.04	-0.66	7.01	-6.93	-2.62
EFP2	-1.63	-0.25	7.32	-7.16	-1.71
(Py)2 P2a(+) SAPT2	-1.78	-0.66	5.90	-6.46	-2.99
EFP2	-0.56	-0.33	5.64	-6.68	-1.93
(Py)2 P2a(-) SAPT2	-3.54	-0.79	7.93	-7.30	-3.70
EFP2	-3.01	-0.28	8.80	-7.48	-1.98
(Py)2 P2b SAPT2	-3.26	-0.70	6.85	-6.89	-4.00
EFP2	-2.98	-0.26	7.08	-7.13	-3.30

Hydrogen bonded. EFP2 agrees well with the estimated CCSD(T)/aug-cc-pVTZ results of ref. 9 for the hydrogen-bonded dimer. The H-bonded dimer is illustrated in Figure 4.4, and the potential energy curves are shown in Figure 4.8. For potential energies between 5.4 Å and 10 Å, the maximum energy difference between the EFP2 and CCSD(T) values is

0.50 kcal/mol (Table 4.2). The root mean square deviation between the two PECs is 0.33 kcal/mol (Table 4.2). Est. CCSD(T)/aug-cc-pVTZ finds the potential minimum for the hydrogen-bonded dimer to be -3.56 kcal/mol at an intermonomer separation (ring center to ring center distance) of 5.8 Å, corresponding to hydrogen bond distances of 2.5 Å. EFP2 finds the minimum to be -3.3 kcal/mol at the same intermonomer separation. At this energy minimum, the Coulomb term is -3.62 kcal/mol, the exchange repulsion 3.95 kcal/mol, the polarization is -0.64 kcal/mol, and the dispersion is -2.99 kcal/mol with EFP2. The magnitude of the exchange-repulsion exceeds that of the Coulomb term by 0.3 kcal/mol. Thus, although the hydrogen bonded configuration exhibits the smallest dispersion interaction in terms of magnitude (compared to the dispersion interaction of the sandwich and T-shaped complexes given in Tables 4.3-4.4), the dispersion energy is necessary for the overall binding of this complex.

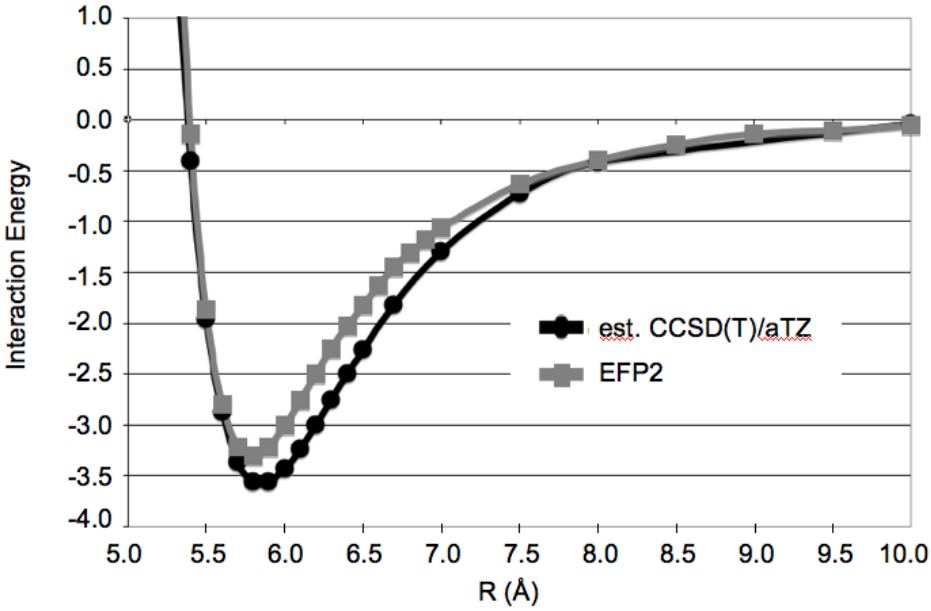


Fig. 4.8. Est. CCSD(T)/aug-cc-pVTZ[9] and EFP2 potential energy curves for the hydrogen bonded pyridine dimer. Interaction energy is in kcal/mol.

Monte Carlo/Simulated Annealing (MC/SA). The primary focus of this investigation has been on the pyridine and benzene-pyridine structures that have been most commonly studied by high-level electronic structure methods. However, the accuracy and computational efficiency of the EFP method suggests that an extensive search for other

minima should be carried out. Consequently, a series of Monte Carlo/simulated annealing (MC/SA) searches were performed.

Important structural motifs found with the EFP2 MC/SA simulations for the benzene-pyridine and pyridine dimers are shown in Figs. 4.9 and 4.10, respectively. Except where noted otherwise, all structures have positive definite Hessians (matrices of second derivatives of the energy with respect to the geometry), indicating that they are minima on the potential energy surface. Very tight gradient convergent criteria were employed for the geometry optimizations. Interaction energies for these structures and the intermonomer distances R_1 and R_2 indicated in Figs. 4.9 and 4.10 appear in Table 4.7. R_1 is the distance between ring centers for all geometries shown. For the lowest energy structures (4.9A, 4.10A), R_2 is the distance between the pyridine heteroatom and the nearest carbon-carbon bond midpoint on the other monomer. The definition of R_2 for other geometries is given in Figs. 4.9 and 4.10.

The lowest energy geometries for Bz-Py and $(Py)_2$ are very similar (Figs. 4.9A and 4.10A). Both structures are found to involve the negatively charged pyridine heteroatom interacting with the positive hydrocarbon backbone of the second monomer. The distance from the pyridine nitrogen to the nearest C-C bond midpoint on the other monomer (R_2) is 3.09 Å for Bz-Py and 3.06 Å for $(Py)_2$. In the Bz-Py lowest energy structure, a line bisecting the pyridine ring through the heteroatom would form a 155° angle with the line given by R. The $(Py)_2$ lowest energy structure has a similar configuration, in which this angle is 156°. A Bz-Py structure in which this angle is 180° (Fig. 4.9B) was found with EFP2 MC/SA (R = 3.10 Å). This structure, which is 0.43 kcal/mol higher in energy than the minimum energy structure, has one imaginary frequency (23.83 cm^{-1}), indicating that the structure in Fig. 4.9B is a transition state between two symmetry-equivalent minima of the Bz-Py dimer. No structure similar to Fig. 4.9B was found by the MC/SA simulations for the $(Py)_2$ dimer.

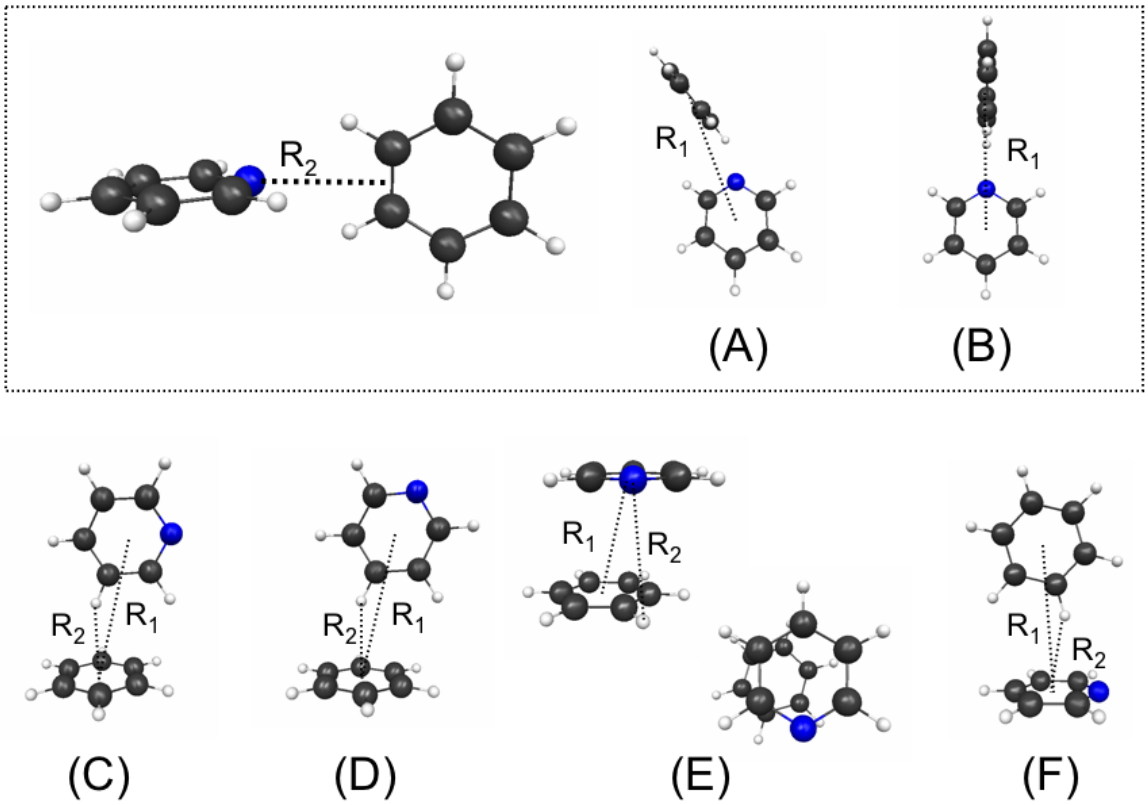


Fig. 4.9. Selected structures found with EFP2 Monte Carlo/simulated annealing (MC/SA) for the benzene-pyridine dimer, in order of strongest to weakest interaction energy. (A) is the lowest energy structure observed. (B) is a transition state between two symmetry-equivalent structures like (A). (C), (D), and (F) are T-like structures, while (E) is a twisted parallel-displaced-type structure. Two views of (E) are shown to illustrate the relative orientation of the monomers. R_1 indicates the distance between monomer centers. For (A) and (B), R_2 indicates the distance between the pyridine heteroatom and the midpoint of the nearest benzene carbon-carbon bond. For (C), (D), and (F), R_2 is the distance from a ring center of one monomer to the nearest hydrogen atom on the other monomer. For (E), R_2 is the distance from the pyridine heteroatom to the nearest benzene hydrogen. Distances and interaction energies of all complexes appear in Table 4.7, while interaction energy components of global minimum structure (A) appear in Table 4.8.

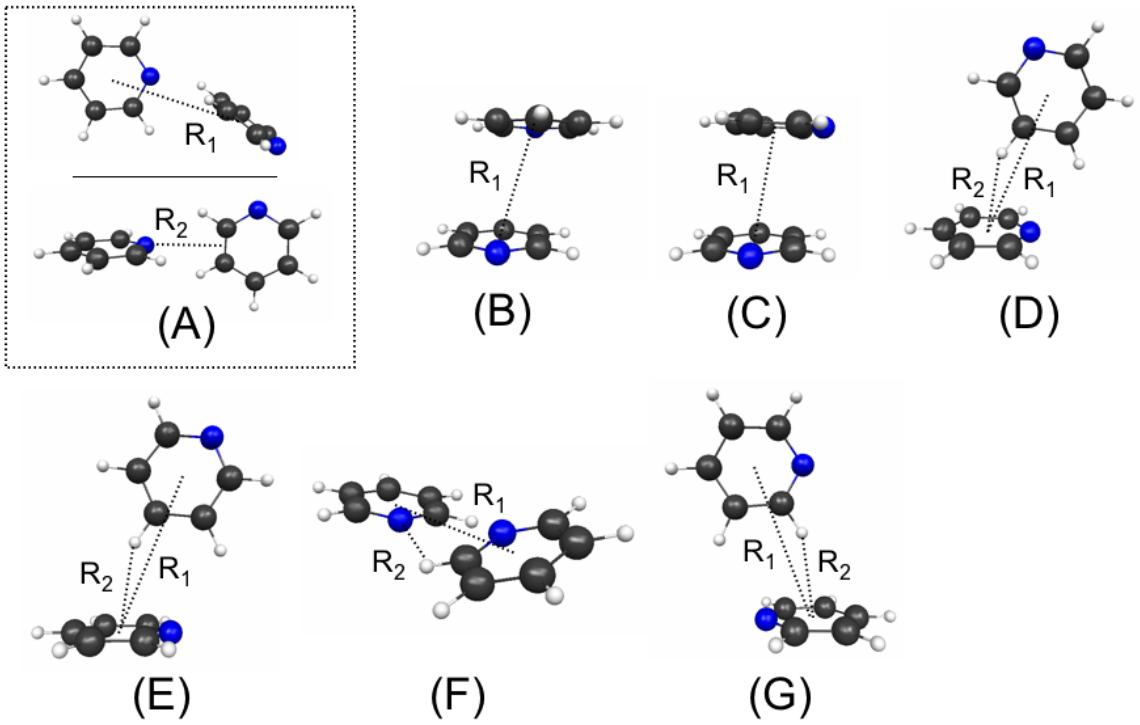


Fig. 4.10. Selected structures found with EFP2 Monte Carlo/simulated annealing (MC/SA) for the pyridine dimer, in order of strongest to weakest interaction energy. (A) is the lowest energy structure observed. (B) and (C) are parallel displaced type structures. (D), (E), and (G) are T-like structures. (F) is a nonplanar hydrogen bonded structure. R_1 indicates the distance between monomer centers. For (A), R_2 indicates the distance between the pyridine heteroatom and the midpoint of the nearest benzene carbon-carbon bond. For (D), (E), and (G), R_2 is the distance from a ring center of one monomer to the nearest hydrogen atom on the other monomer. For (F), R_2 indicates the length of the hydrogen bond. For (B) and (C), an R_2 value is not defined. Distances and interaction energies of all complexes appear in Table 4.7, while interaction energy components of global minimum structure (A) appear in Table 4.8.

Table 4.7. EFP2 interaction energies and intermonomer distances of selected Bz-Py and $(Py)_2$ structures found with EFP2 Monte Carlo/simulated annealing (MC/SA). Interaction energies ΔE (kcal/mol) and intermonomer distances R_1 and R_2 (Å) of the structures shown in Fig. 4.9-4.10. R_1 is the distance between monomer centers in all cases. R_2 is defined in Fig. 4.9-4.10 for each structure.

	R_1	R_2	ΔE
Bz-Py			
A	5.55	3.09	-4.68
B	5.71	3.10	-4.25
C	4.96	2.64	-3.63
D	4.98	2.65	-3.60
E	3.87	3.64	-3.46
F	4.94	2.56	-3.19
(Py)2			
A	5.53	3.06	-5.39
B	3.80		-4.23
C	3.79		-4.11
D	4.93	2.61	-4.10
E	4.96	2.63	-3.83
F	5.55	2.44	-3.70
G	4.92	2.65	-3.67

The EFP2 interaction energy components for the lowest energy geometries found with the MC/SA simulations (Fig. 4.9A, 4.10A) are given in Table 4.8. Both structures have exchange-repulsion terms of approximately 6.5 kcal/mol, and have correspondingly large Coulomb interaction terms as well: -6.10 kcal/mol for Bz-Py MC/SA, -6.85 kcal/mol for $(Py)_2$ MC/SA (Table 7). Due to the angle between the monomers (described above), which brings the ortho hydrogen of pyridine closer to the π cloud of the other monomer, the dispersion interaction energy is also large for these structures. This creates a "π -hydrogen bond" effect similar to that which occurs in the T-shaped dimers. The dispersion energy of the Bz-Py global minimum structure 9A (-4.33 kcal/mol, Table 4.7) is approximately the same as that of Bz-Py T1 (-4.34 kcal/mol) or Bz-Py T2 (-4.35 kcal/mol, Table 4.4). The dispersion energy of the $(Py)_2$ global minimum structure is a very similar -4.34 kcal/mol (Table 4.7), greater than that of any $(Py)_2$ T-shaped structures except $(Py)_2$ T2 (-4.70 kcal/mol, Table 4.4). Thus, while the Coulomb term is the single largest attractive energy term for Bz-Py 4.9A, the structure would not be bound without the dispersion interaction. The $(Py)_2$ lowest energy structure 10A is very similar to that found for Bz-Py. Overall, $(Py)_2$ 4.10A is 0.71 kcal/mol more strongly bound than Bz-Py 4.9A, due mainly to a larger

Coulomb attraction. While the $(Py)_2$ 10A structure would still be weakly bound without the dispersion interaction, the dispersion interaction does contribute -4.34 kcal/mol to the binding energy.

Table 4.8. EFP2 energy contributions to the global minimum benzene-pyridine and pyridine dimer structures found with EFP2 Monte Carlo/simulated annealing. Energies in kcal/mol.

	Coulomb	Pol	Exch-Rep	Dispersion	Total
Bz-Py MC/SA	-6.10	-0.68	6.43	-4.33	-4.68
(Py)2 MC/SA	-6.85	-0.79	6.59	-4.34	-5.39

For the Bz-Py dimer, in addition to the lowest energy structure 4.9A and the transition state 9B, structures resembling T-shaped (Fig. 4.9C,D,F) and parallel displaced (Fig. 4.9E) motifs were also found. The T-like structure 4.9C resembles Bz-Py T4, although the ring center of pyridine does not lie directly over the ring center of benzene as in the artificially constrained Bz-Py T4. In structure 4.9C, the meta carbon of pyridine is closest to the benzene ring center. The T-like structure 4.9D is similar to 4.9C, but the para carbon of pyridine lies most directly over the benzene ring center. Structures 4.9C and 4.9D are nearly degenerate in energy, differing by only 0.03 kcal/mol. They are, respectively, 1.05 kcal/mol and 1.08 kcal/mol higher in energy than the lowest energy structure 4.9A (Table 4.7). The parallel displaced structure 4.9E most closely resembles Bz-Py P1b, although the monomers in 4.9E are rotated compared to their orientation in Bz-Py P1b. This dimer configuration is 1.22 kcal/mol less strongly bound than the lowest energy configuration 4.9A (Table 4.7). Finally, the T-like structure 4.9F is most similar to Bz-Py T3, in that a benzene hydrogen atom lies over the pyridine ring. However, compared to its orientation in Bz-Py T3, the benzene in 9F is rotated 90° relative to pyridine. Also, as with the other T-like structures, the ring centers do not lie directly over one another. This structure is 1.49 kcal/mol higher in energy than the minimum energy structure 4.9A (Table 4.7).

Due to the lower symmetry of pyridine compared to benzene, the potential energy surface of $(Py)_2$ is even more diverse than that of Bz-Py. A selection of unique structures found with MC/SA are shown in Fig. 4.10, with structure 4.10A being the lowest in energy. Structures 4.10B and 4.10C are parallel displaced structures with different orientations of the pyridine monomers. The more energetically favorable (4.10B) has antiparallel pyridines,

similar to $(Py)_2$ P2b, while in the other (4.10C), the pyridine heteroatoms are offset by 90°. These structures differ in energy by 0.12 kcal/mol, with the antiparallel structure (4.10B) 1.16 kcal/mol higher in energy than the minimum energy configuration 10A (Table 4.7). Structures 4.10D, 4.10E, and 4.10G are T-shaped type, although the monomers are offset: the center of the "top" monomer does not lie directly above the ring plane of the other. Structure 10G most closely resembles $(Py)_2$ T3, with the "top" monomer displaced over the heteroatom of the other. The geometries of 4.10D and 4.10E do not directly correspond to any of the constrained T-shaped structures examined previously in this study; in addition to the offset monomer centers, the "top" pyridine is rotated ~60° compared to its orientation in otherwise similar constrained structures (e.g. $(Py)_2$ T3). The T-shaped structure 4.10D, at -4.10 kcal/mol, is nearly isoenergetic with the parallel displaced structure 4.10C, at -4.11 kcal/mol (Table 4.7). Configurations 4.10D, 4.10E, and 4.10G are respectively 1.29 kcal/mol, 1.56 kcal/mol, and 1.72 kcal/mol higher in energy than the minimum energy configuration 4.10A (Table 4.7). Fig. 4.10F has a nonplanar hydrogen bonded structure. A planar H-bonded geometry was not found with the MC/SA simulations, and an EFP2 geometry optimization beginning from the geometry of the most energetically favorable planar H-bonded structure examined previously in this study yields a second-order saddle point. The nonplanar H-bonded structure (4.10F) lies 1.69 kcal/mol above the minimum energy geometry (Table 4.7).

The offset T-shaped (Fig. 4.9C, 4.9D, 4.9F, 4.10D, 4.10E, 4.10G) and parallel displaced (4.9E, 4.10B, 4.10C) structures found with the EFP2 MC/SA simulations have analogues on the benzene dimer potential energy surface. A previous EFP2 MC/SA study[7] that included the benzene dimer found structures similar to those discussed above. For the benzene dimer, the EFP2 lowest energy structure is offset T-shaped, with a monomer center-to-center distance (R_1) of 5.1 Å, and an interaction energy of -2.80 kcal/mol. A parallel displaced structure with R_1 = 4.0 Å has an interaction energy of -2.41 kcal/mol. Finally, an edge-to-edge configuration of the benzene dimer similar to 4.9B, with R_1 = 6.2 Å, has an interaction energy of -1.53 kcal/mol. These EFP2 results are similar to the findings of a previous study[23] examining 491 configurations of the benzene dimer. Energies computed with MP2/aug-cc-pVTZ + ΔCCSD(T) correction of similar offset T-shaped, parallel displaced, and edge-to-edge structures are reported as -2.795 kcal/mol, -2.699 kcal/mol, and -

1.805 kcal/mol, respectively; these structures are also reported to be true minima, rather than saddle points.

The EFP2 MC/SA results can also be compared with results from a DFT-D study[10] of the pyridine dimer potential energy surface. A structure similar to 10E and a planar hydrogen bonded structure were found with DFT-D.[10] The DFT-D structure had a hydrogen-to-ring-center distance R_2 of 2.66 Å [ref. 10] compared to 2.63 Å found by EFP2 (Table 4.7). The EFP2 interaction energy for this structure is -3.8 kcal/mol (Table 4.7), compared to -3.1 kcal/mol with SCS-MP2, -4.2 kcal/mol with MP2, and -3.0 kcal/mol with DFT-D. As noted above, EFP2 finds a planar H-bonded structure (with hydrogen bond lengths R_2 = 2.44 Å v. 2.54 Å with the DFT-D[10] optimization) that is a second-order saddle point. The authors of the DFT-D study performed a full geometry optimization and noted that use of constrained monomer geometries altered the results of the study.[10] It is possible that the lack of planarity of the optimum EFP2 H-bonded structure results from the use of rigid monomers; however, no mention is made of Hessian calculations in ref. 10, so it is not known whether the H-bonded structure reported there is a true minimum. An edge-to-edge structure like the EFP2 global minimum 4.10A is not reported in the DFT-D study.[10]

Cartesian coordinates and interaction energy components for the MC/SA structures described above are available online as Supplementary Information.

Conclusions

For dimers of benzene, pyridine, and benzene-pyridine, the general effective fragment potential method (EFP2) is in good agreement with the high-level CCSD(T), MP2, and SAPT2 results of Hohenstein and Sherrill[9]. Consistent with previous studies on benzene dimers[5] and substituted benzene dimers[7], EFP2 is found to slightly overestimate binding energies and intermonomer separations in sandwich structures and T-shaped structures other than those of type T2. EFP2 underestimates the Coulomb attraction in Bz-Py T2 and $(Py)_2$ T2, resulting in no bound state for the former and a weakly bound minimum for the latter. Slight differences in the energy order of the T-shaped dimer curves arise because EFP2 overbinds the T-shaped benzene dimer slightly more than other T-shaped structures.

However, EFP2 underbinds parallel displaced benzene dimers more than it underbinds other parallel displaced structures.

Overall, the differences between EFP2 and CCSD(T) or MP2 are small. The root mean square deviation (RMSD) between EFP2 and CCSD(T) potential energy curves for the sandwich, T-shaped, and hydrogen bonded dimers is 0.49 kcal/mol, with a range from 0.31 kcal/mol (Bz-Py T3) to 0.66 kcal/mol ($(Py)_2$ S2). The maximum unsigned difference between EFP2 and CCSD(T) energies is 1.08 kcal/mol among sandwich dimers ($(Py)_2$ S2), while it is 1.97 among T-shaped dimers (Bz-Py T1), for which EFP2 tends to predict a much more rapid increase in potential energy at short intermonomer separations than CCSD(T). The RMSD between the EFP2 and SCS-MP2 curves for the parallel displaced dimers is 0.52 kcal/mol, with a maximum unsigned difference of 1.31 kcal/mol (for both types of parallel displaced benzene dimers). On average, the EFP2 potential energy curves for all dimers examined differ from the corresponding high-level *ab initio* curves by just 0.50 kcal/mol. EFP2 also provides an energy decomposition that compares well with SAPT2 results, making it an attractive method for use with aromatic heterocycles.

Dispersion and Coulomb interactions are the principal attractive forces in the dimers studied. Even in the hydrogen-bonded pyridine dimer, which is expected to be stabilized primarily by Coulomb attraction, the exchange-repulsion is of a similar magnitude. A similar observation is made for the lowest energy structures found with Monte Carlo/simulated annealing. While both the Bz-Py and $(Py)_2$ EFP2 MC/SA minimum-energy geometries involve direct interaction between the negatively-charged heteroatom of pyridine and the positive carbon backbone of the other monomer, the exchange-repulsion exceeds (in the case of Bz-Py) or nearly exceeds (in the case of $(Py)_2$) the Coulomb attraction. An accurate depiction of dispersion energy is necessary to describe the binding of these dimers.

Acknowledgements

This work was supported by a grant (to MSG) from the Air Force Office of Scientific Research, and by a NSF CAREER and ACS PRF grants (to LVS). The authors are grateful to Ed Hohenstein and Professor David Sherrill for providing high-level *ab initio* data for the

benzene and pyridine dimers. The authors also gratefully acknowledge Iowa State University and Purdue University.

Supporting Information

Cartesian coordinates of the minimum energy benzene-pyridine and pyridine dimer structures are available online, along with the coordinates and interaction energy components of several additional low-lying geometries found with EFP2 Monte Carlo/simulated annealing. This material is available free of charge via the Internet at http://pubs.acs.org.

References

1) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23. Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837.
2) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
3) Lerman, L. S. *J. Mol. Biol.* **1961**, *3*, 18. Brana, M. F.; Cacho, M.; Gradillas, A.; Pascual-Teresa, B.; Ramos A. *Curr. Pharm. Des.* **2001**, *7*, 1745.
4) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevents, W. J. *J. Phys. Chem. A* **2001**, *105*, 293. Jensen, J. H.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.
5) Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.
6) Gordon, M. S.; Slipchenko, L. V.; Li, H.; Jensen, J. H. *Annual Reports in Comp. Chem.* **2007**, *3*, 177.
7) Smith, T.; Slipchenko, L. V.; Gordon, M. S. *J. Phys. Chem. A* **2008**, *112*, 5286
8) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.
9) Hohenstein, E. G.; Sherrill, C D. *J. Phys. Chem A* **2009**, *113*, 878.
10) Piacenza, M; Grimme, S. *ChemPhysChem* **2005**, *6*, 1554.
11) Mishra, B. K.; Sathyamurthy, N. *J. Phys. Chem. A* **2005**, *109*, 6.
12) Möller, C; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
13) A. D. McLean and G. S. Chandler, J. Chem. Phys. **1980**, *72*, 5639.
14) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem*. **1993**, *14*, 1347.
15) Stone, A. J. *The Theory of Intermolecular Forces.* Oxford University Press: Oxford, **1996**.
16) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128.
17) Jenson, J. H.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.
18) Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *103*, 379. Amos, R. D.; Handy, N. C.;

Knowles, P. J.; Rice, J. E.; Stone, A. J. *J. Phys. Chem.* **1985**, *89*, 2186.
19) Li, H.; Gordon, M. S.; Jensen, J. H. *J. Chem. Phys.* **2006**, *124*, 214108.
20) Slipchenko, L. V.; Gordon, M. S. *Mol. Phys.* **2009**, *107*, 999.
21) Gauss, J.; Stanton, J. F. *J. Phys. Chem. A* **2000**, *104*, 2865
22) Innes, K. K.; Ross, I. G.; Moomaw, W. R. *J. Mol. Spectrosc*. **1988**, *132*, 492.
23) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A,* **2006**, *110* (34), 10345.
24) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007. Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
25) Haliker, A.; Klopper, W.; Helgaker, T.; Jorgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157.
26) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
27) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
28) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev*. **1994**, *94*, 1887. Williams, H. L.; Szalewicz, K.; Jeziorski, B.; Moszynski, R.; Rybak, S. *J. Chem. Phys*. **1993**, *98*, 1279.
29) Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys.* **2000**, *112*, 2063.
30) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200; Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887.

CHAPTER 5: INTERACTIONS BETWEEN PAIRED DNA NUCLEOTIDE BASES MODELED WITH THE EFFECTIVE FRAGMENT POTENTIAL METHOD

To be submitted to the *Journal of Physical Chemistry A*

Q. A. Smith, L. V. Slipchenko, and M. S. Gordon

Abstract

Hydrogen bonded and stacked structures of adenine-thymine and guanine-cytosine nucleotide base pairs, along with their methylated analogues, are examined with the *ab inito*-based general effective fragment potential (EFP2) method. A comparison with coupled cluster with singles, doubles, and perturbative triples (CCSD(T)) energies is presented, along with an EFP2 energy decomposition to illustrate the components of the interaction energy.

Introduction

Accurately modeling complexes as large as paired DNA nucleotide bases with *ab initio* methods remains difficult and computationally expensive. The most complete set of high-level *ab initio* calculations on these systems is that of Hobza et al[1], performed at the estimated coupled cluster with singles, doubles, and perturbative triples (CCSD(T))[2] level of theory. Hobza's work provides a benchmark with which to test the accuracy of other methods. The general effective fragment potential (EFP2) method,[3] is a fast, *ab initio*-based method that has shown success in modeling dimers of benzene,[4] benzene-water,[5] substituted benzenes,[6] and pyridine,[7] systems with properties similar to DNA. A preliminary report employed some EFP2 features to examine stacked adenine-adenine and thymine-thymine dimers, as well as hydrogen-bonded adenine-thymine.[8] This work expands upon the previous study, by using the complete EFP2 method to model both the fundamental hydrogen bonded and stacked DNA nucleotide base pairs of guanine-cytosine and methylated adenine-thymine and guanine-cytosine.

The adenine-thymine (A…T) and guanine-cytosine (G…C) Watson-Crick (WC) or hydrogen bonded structures are the canonical nucleotide base pairs. Their methylated analogues are denoted mA…mT (9-methyladenine–1-methylthymine) and mG…mC (9-methylguanine–1-methylcytosine). Multiple hydrogen bonds give these complexes large interaction energies that are dominated by Coulomb forces.[1] G…C WC and mG…mC WC are stabilized by three hydrogen bonds, denoted R_1, R_2, and R_3 in Figs. 5.1A and 5.2A. A…T WC and mA…mT WC have two hydrogen bonds, denoted R_1 and R_2 in Figs. 5.3A and 5.4A. When paired in DNA, the oxygen at thymine carbon C2 and the hydrogen at adenine carbon C2 (the intermonomer separation labeled R_3 in Figs. 5.3A and 5.4A) are too widely separated and nonlinear to contribute to hydrogen bonding, although this distance is shown for descriptive purposes.

The stacked nucleotide base structures are more difficult to describe than the hydrogen bonded species without resorting to very high levels of theory. The difficulty in modeling these structures arises because of the inability of most methods, even the simpler electronic structure methods (e.g., Hartree-Fock (HF) and the most commonly used density functional theory (DFT) functionals), to accurately portray the dispersion energy.[9] Second-order perturbation theory[10] (MP2) is, in addition, known to overestimate the binding energy of stacked nucleotide bases by about 20%, relative to the more reliable (and more costly) coupled cluster method with single, double and perturbative triple (CCSD(T)) excitations.[1,11] Indeed, the ΔCCSD(T) correction term (the difference between CCSD(T) and MP2 energies) is positive for the stacked nucleotide bases, in contrast to the hydrogen-bonded structures for which the correction is negative.[1] DFT often fails to find any bound nucleotide base pairs with a stacked motif,[12] although DFT can be made to perform better when dispersion is explicitly introduced through parameterization or hybrid methods.[13-15]

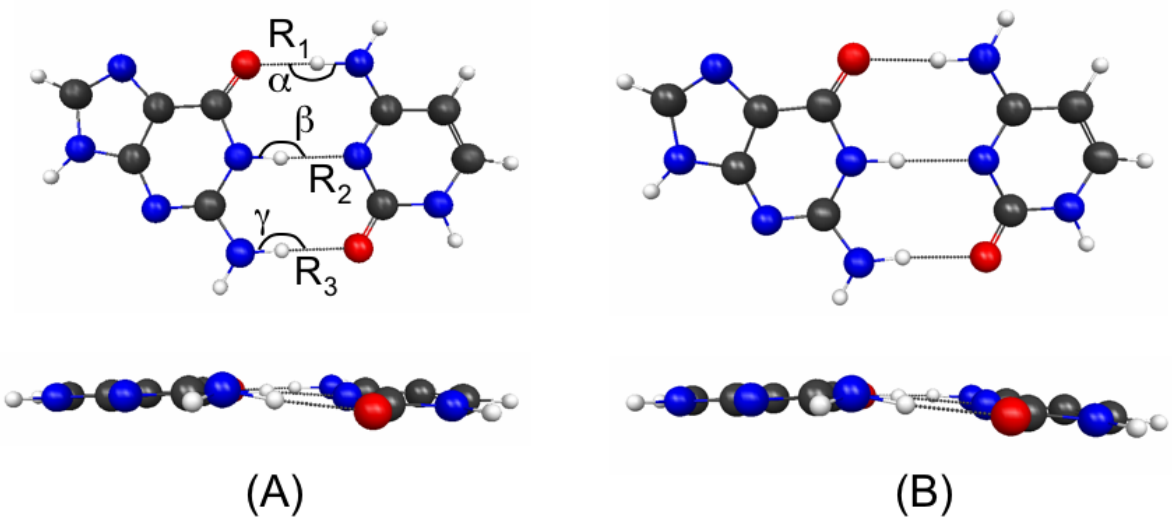


Fig. 5.1. Top and side views of the guanine-cytosine hydrogen bonded complex. (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref. 1. (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref. 1. Hydrogen bond lengths R_1, R_2, and R_3, and angles α, β, and γ (chosen to show the linearity of each hydrogen bond) are given in Table 5.1 for each complex.

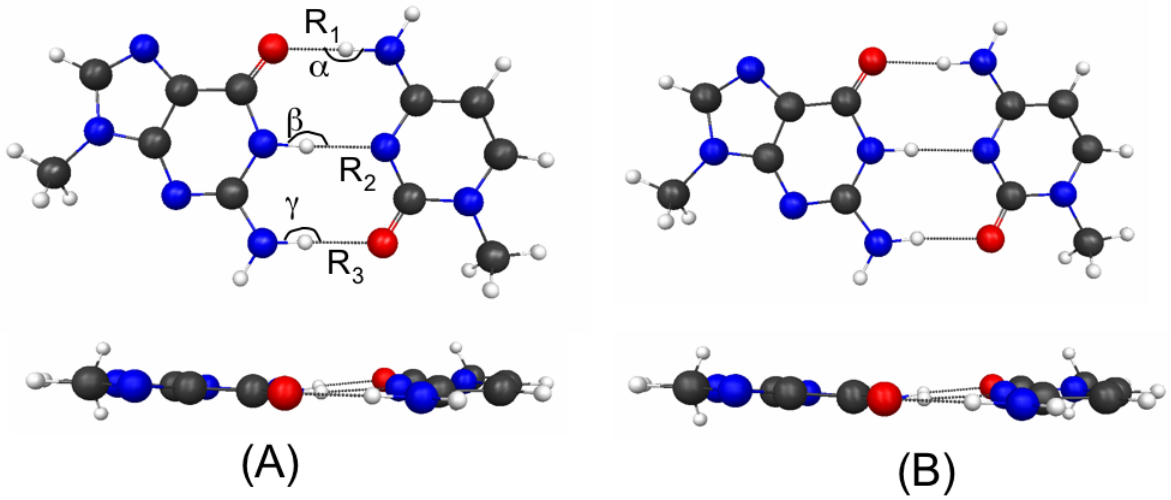


Fig. 5.2. Top and side views of the methylated guanine-cytosine hydrogen bonded complex. (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref. 1. (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref. 1. Hydrogen bond lengths R_1, R_2, and R_3 and angles α, β, and γ (chosen to show the linearity of each hydrogen bond) are given in Table 5.1 for each complex.

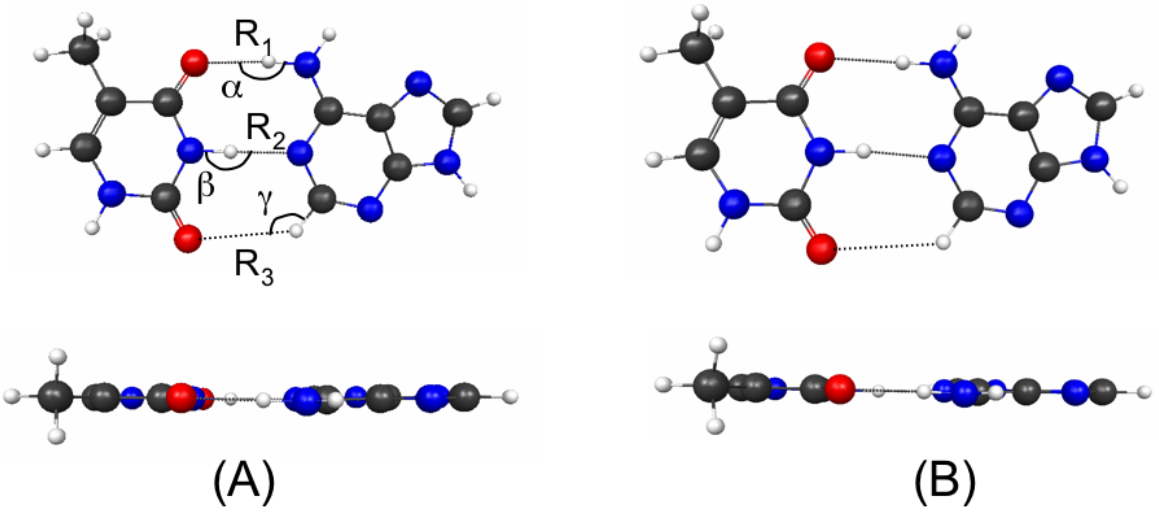


Fig. 5.3. Top and side views of the adenine-thymine hydrogen bonded complex. (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref. 1. (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref. 1. Intermonomer distances R_1, R_2, and R_3 and angles α, β, and γ (chosen to show the linearity of each hydrogen bond) are given in Table 5.1 for each complex.

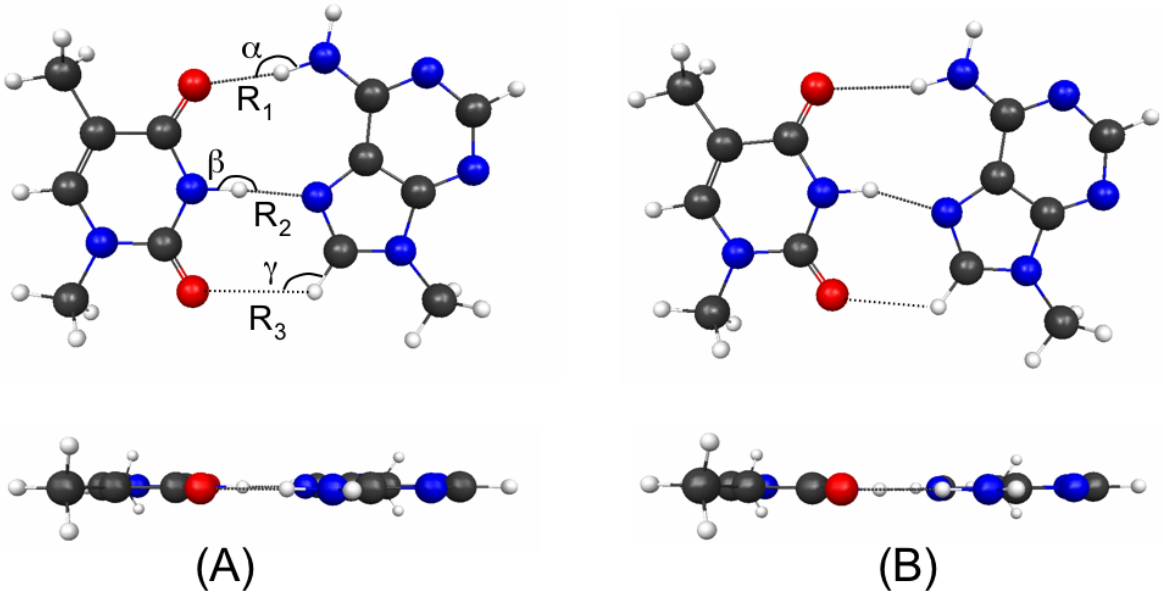


Fig. 5.4. Top and side views of the methylated adenine-thymine hydrogen bonded complex. (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref. 1. (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref. 1. Intermonomer distances R_1, R_2, and R_3 and angles α, β, and γ (chosen to show the linearity of each hydrogen bond) are given in Table 5.1 for each complex.

Methods

The general effective fragment potential (EFP2) method is coded in the GAMESS (General Atomic and Molecular Electronic Structure System)[16] quantum chemistry software package, which was used for all calculations in this study. The EFP2 total interaction energy is decomposed into Coulomb, exchange-repulsion, polarization (induction), dispersion, and charge transfer terms. The Coulomb interaction is calculated using Stone's distributed multipolar expansion,[17] carried out through octopole moments. The analytical distributed multipolar analysis (DMA) was used in this study, although a numerical DMA[18] is also available. Charge penetration, required when fragments approach each other closely, is modeled by an exponential damping function that multiplies the distributed multipoles, up through the dipole-quadrupole term, but not the quadrupole-quadrupole or higher order terms.[19] Exchange-repulsion is derived as an expansion in the intermolecular overlap, truncated at the quadratic term.[20] Polarization is represented by the sum of localized molecular orbital (LMO) polarizability tensors. Polarizable points are located at the LMO centroids, corresponding to the bonds and lone pairs of the molecule. These LMO polarizabilities are determined using the coupled-perturbed Hartree-Fock equations.[21] The dispersion interaction is expressed in terms of imaginary frequency-dependent polarizabilities, with an explicitly derived C_6 coefficient and an estimated C_8 coefficient.[21] Charge transfer, important in systems with charged or highly polar molecules, is computed from a perturbative treatment of the interaction between occupied orbitals on one fragment and virtual orbitals on a second fragment.[22] Charge transfer is not included in the present study because a previous study[22] demonstrated that this term is very small for most neutral molecules.

Geometries. Two sets of adenine-thymine (A…T) and guanine-cytosine (G…C) stacked and hydrogen bonded geometries were examined with the EFP2 method, along with two sets of the corresponding methylated structures: 9-methyladenine–1-methylthymine (mA…mT) and 9-methylguanine–1-methylcytosine (mG…mC). All EFP2 potentials were generated with the analytic DMA[23] and the 6-311++G(3df,2p) basis set. EFP2 fragments are rigid, having fixed internal geometries.

Set 1. The first set of structures was taken directly from the JSCH-2005 benchmark data[1]. The stacked structures chosen for the present study are denoted A…T S, G…C S, mA…mT S, and mG…mC S in JSCH-2005; the hydrogen bonded structures are denoted A…T WC [Watson-Crick], G…C WC, mA…mT WC, and mG…mC WC. The geometries for all of these structures were obtained through optimization with the RI-MP2/cc-pVTZ method.[1] To analyze these structures with the EFP2 method, EFP potentials were generated separately for each nucleotide base. These EFP2 nucleotide bases were then recombined to produce paired geometries identical to the JSCH-2005 structures. This set tests the ability of the EFP2 method to reproduce estimated CCSD(T) interaction energies[1].

Set 2. The second set of nucleotide base geometries began with the same nucleotide bases and EFP2 potentials described in Set 1. However, in this case, instead of constraining the nucleotide bases to the paired geometries found with RI-MP2 in ref. 1, an EFP2 geometry optimization was performed. This set tests the ability of the EFP2 method to reproduce paired geometries.

Interaction energies. The estimated CCSD(T) interaction energies are taken from ref. 1. In that work, RI-MP2 energies were counterpoise corrected for basis set superposition error (BSSE)[24] and extrapolated to the complete basis set (CBS) limit by the two-point extrapolation scheme of Helgaker *et al*.[25] Small-basis CCSD(T) single point energy calculations were performed at the RI-MP2 optimized geometries; the difference between the small-basis CCSD(T) energy and the small-basis MP2 energy is the CCSD(T) energy correction term, ΔCCSD(T). The final estimated CCSD(T)/CBS interaction energy is given by $\Delta E_{CBS}^{CCSD(T)} = \Delta E_{CBS}^{MP2} + \Delta CCSD(T)$, where $\Delta CCSD(T) = \left(\Delta E^{CCSD(T)} - \Delta E^{MP2}\right)_{small-basis}$.

Components of the interaction energy. A full symmetry adapted perturbation theory (SAPT) study of paired nucleotide bases has not been undertaken, to the authors' knowledge. Previous studies combining SAPT with density functional theory (DFT) exist,[14,15] but a DFT-based treatment of nucleotide base interactions may be insufficient, because DFT functionals often perform poorly at modeling dispersion.[12] Thus, no definitive data exists with which to compare the EFP2 interaction energy components. EFP2 has previously been shown to agree closely with SAPT results in systems similar to the paired

nucleotide bases examined here (e.g. dimers of benzene,[4] substituted benzenes,[6] and pyridine[7]).

Results and Discussion

Hydrogen bonded structures. EFP2 geometry optimizations (Set 2) accurately predict the structures of the hydrogen bonded complexes. The lengths of the hydrogen bonds corresponding to R_1, R_2, and R_3 (Figs. 5.1 and 5.2) or to R_1 and R_2 (Figs. 5.3 and 5.4) are given in Table 5.1. The intermonomer separation R_3 is also given in Figs. 5.3 and 5.4, although R_3 is not a true hydrogen bond. The angles denoted by Greek letters in Figs. 5.1 - 5.4 are chosen to show the linearity of each hydrogen bond. These values are reported in Table 5.1.

Table 5.1. Geometries of hydrogen bonded complexes. Lengths (in Å) and angles (in degrees) of the hydrogen bonds in the Watson-Crick (WC) complexes of Figs. 1 - 4. "Opt type" refers to the level of theory used for the geometry optimization, either RI-MP2 [ref. 1] or EFP2. Lengths are indicated by A, B, and C in Figs. 5.1 and 5.2 for guanine-cytosine (G…C) and methylated guanine-cytosine (mG…mC) and by A and B in Figs. 5.3 and 5.4 for adenine-thymine (A…T) and methylated adenine-thymine (mA…mT). For A…T and mA…mT, intermonomer separation C is also given, although it is not a true hydrogen bond.

	opt type	R_1	R_2	R_3	α	β	γ
G…C WC	MP2	1.76	1.91	1.92	178.5	175.1	175.6
	EFP2	1.87	1.98	1.95	175.3	175.3	172.4
mG…mC WC	MP2	1.88	1.87	1.73	175.3	176.0	178.7
	EFP2	1.83	1.97	1.95	176.0	175.3	171.6
A…T WC	MP2	1.93	1.82	2.83	173.6	179.1	132.7
	EFP2	2.01	1.93	2.94	178.5	173.5	127.5
mA…mT WC	MP2	1.93	1.75	2.72	169.3	175.7	120.8
	EFP2	2.23	1.78	2.39	166.2	159.9	119.1

The EFP2 G…C WC hydrogen bond lengths differ from the RI-MP2/cc-pVTZ[1] lengths by 0.03 - 0.11 Å (Table 5.1). The EFP2 A…T WC hydrogen bond lengths are 0.08 - 0.11 Å longer than those found with RI-MP2[1]. The nonmethylated EFP2 structures depicted in Figs. 5.1B and 5.2B are very similar to the RI-MP2[1] structures of Figs. 5.1A and 5.2A, respectively. Structural differences are greater between the EFP2-optimized methylated WC

geometries compared to the corresponding RI-MP2[1] geometries. In EFP2-optimized mG…mC WC, the hydrogen bond R_1 in Fig. 5.2 is 0.05 Å shorter than it is in RI-MP2 optimized[1] mG…mC WC, while R_3 is 0.22 Å longer. In EFP2-optimized mA…mT WC, R_3 in Fig. 5.4 is 0.33 Å shorter than in RI-MP2 optimized[1] mA…mT WC, while R_1 is 0.30 Å longer. This indicates that the methylated nucleotide bases optimized with EFP2 are slightly tilted relative to their orientation in the RI-MP2 optimized[1] configuration.

Table 5.2. Energies of hydrogen bonded complexes. EFP2 energy components (Coulomb, exchange-repulsion, polarization, dispersion) and total interaction energy for the hydrogen bonded nucleotide base complexes shown in Figs. 5.1 - 5.4. "Opt type" refers to the level of theory used for the geometry optimization, either RI-MP2 [ref. 1] or EFP2. EFP2 total energies are compared with the estimated CCSD(T)/CBS interaction energies[1] shown for complexes with RI-MP2 optimized geometries taken from ref. 1. Energies in kcal/mol.

	opt type	EFP2 COULOMB	EFP2 EX-REP	EFP2 POL	EFP2 DISP	EFP2 TOTAL	CCSD(T)
G…C WC	MP2	-41.4	36.4	-14.7	-10.7	-30.4	-32.1
	EFP2	-37.6	28.1	-12.3	-9.4	-31.2	
mG…mC WC	MP2	-44.4	41.4	-16.6	-11.7	-31.3	-31.6
	EFP2	-39.4	30.6	-13.5	-10.0	-32.3	
A…T WC	MP2	-25.3	26.7	-7.2	-8.2	-14.0	-16.9
	EFP2	-22.0	19.0	-5.4	-7.0	-15.3	
mA…mT WC	MP2	-28.8	29.8	-8.1	-9.1	-16.1	-18.2
	EFP2	-27.5	26.3	-6.9	-9.2	-17.3	

The EFP2 total interaction energies (Table 5.2) of the hydrogen bonded nucleotide bases are generally in good agreement with the estimated CCSD(T) interaction energies[1]. EFP2 underbinds the G…C WC 1 complex by 1.7 kcal/mol (~5% of the total binding energy) compared to the estimated CCSD(T) value[1]. When allowed to relax to the slightly (0.03 - 0.11 Å) more widely separated geometry found with an EFP2 optimization, the EFP2 G…C WC total interaction energy differs from the estimated CCSD(T) value[1] by less than 1 kcal/mol. The EFP2 interaction energy of the RI-MP2 optimized[1] methylated guanine-cytosine complex differs from the estimated CCSD(T) value[1] by only 0.3 kcal/mol. The greatest discrepancy between the EFP2 and estimated CCSD(T)[1] interaction energies is in the adenine-thymine complex. EFP2 under-binds the structure at both geometries examined, by 2.9 kcal/mol when constrained to the RI-MP2[1] geometry and by 1.6 kcal/mol when an EFP2 geometry optimization is used. EFP2 also under-binds the RI-MP2 optimized[1] methylated

adenine-thymine structure compared to the estimated CCSD(T) value[1], by 2.1 kcal/mol. Although this discrepancy decreases to less than 1 kcal/mol with the EFP2-optimized structure, the EFP2-optimized structure also differs the most from its corresponding RI-MP2[1] geometry, as discussed above.

The Coulomb interaction is the predominating attractive force in the hydrogen bonded complexes (Table 5.2), although, especially in the case of adenine-thymine and their methylated analogues, the exchange-repulsion is similar in size and opposite in sign. The EFP2 exchange-repulsion term exceeds the Coulomb attraction in the RI-MP2[1] geometries of A…T WC and mA…mT WC. At the EFP2-optimized geometries, the Coulomb magnitude exceeds the exchange-repulsion magnitude by only 3.0 kcal/mol in A…T WC and by 1.2 kcal/mol in mA…mT WC. While the Coulomb term is ~5-10 kcal/mol larger than the exchange-repulsion in all G…C WC and mG…mC WC structures, the total EFP2 interaction energies are –30.4 kcal/mol (RI-MP2 optimized[1] G…C WC) to -32.3 kcal/mol (EFP2-optimized mG…mC WC). The Coulomb + exchange-repulsion accounts for only ~10% (RI-MP2 optimized[1] mG…mC WC) to 30% (EFP2-optimized G…C WC) of the total interaction energy. Thus, polarization and dispersion make significant contributions to the binding of hydrogen bonded nucleotide base pairs.

Stacked structures. EFP2 geometry optimizations produce stacked structures for guanine-cytosine (Fig. 5.5B), methylated guanine-cytosine (Fig. 5.6B), and adenine-thymine (Fig. 5.7B) that are similar to their RI-MP2 optimized[1] counterparts (Figs. 5.5A - 5.7A). The EFP2-optimized structure for methylated adenine-thymine (Fig. 5.8B) is, however, significantly different from that obtained with RI-MP2[1] (Fig. 5.8A); between these two structures, the dihedral angle between nucleotide base ring planes changes by 65°.

The EFP2-optimized stacked guanine-cytosine complex (G…C S) is similar to the RI-MP2 optimized[1] G…C S. The nucleotide bases in EFP2-optimized G…C S are just 0.07 Å more widely separated than in the RI-MP2 optimized[1] G…C S structure (Table 5.3), and they are oriented only slightly differently, with the dihedral angle between the ring planes (Fig. 5.5A) differing by 6.3°. The greatest difference is their relative displacement, indicated by the angle α, which differs by ~10°. EFP2 interaction energies for RI-MP2 optimized[1] G…C S and EFP2-optimized G….C S differ by about 1 kcal/mol (Table 5.4). The EFP2

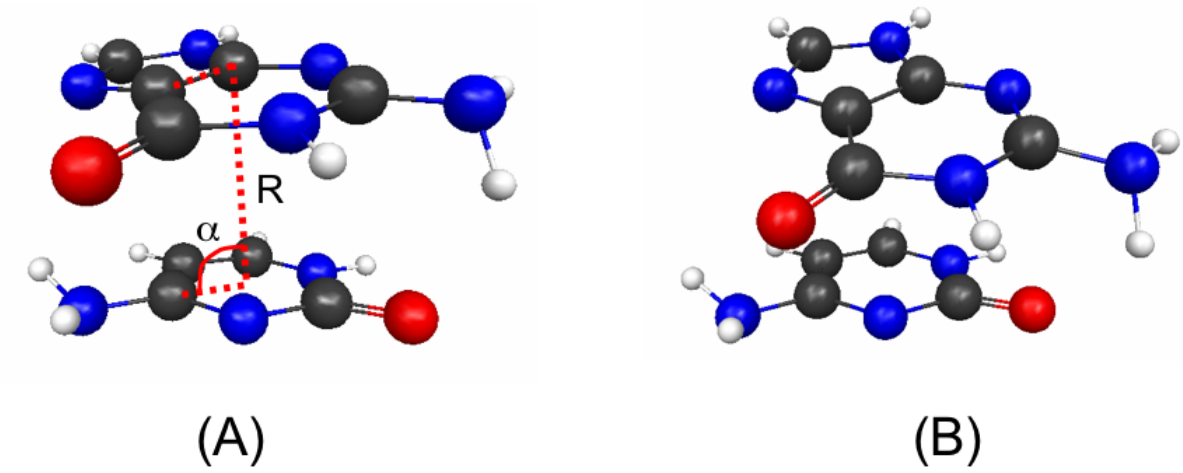


Fig. 5.5. Guanine-cytosine stacked complex. (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref. 1. (B) is the complex obtained by EFP2 optimization with the nucleobase geometries of ref. 1. Intermonomer distance R, angle α (chosen to demonstrate the relative displacement of the nucleotide bases with respect to one another), and the dihedral angle between the planes of the nucleotide bases (dashed red line) are given in Table 5.3 for each complex.

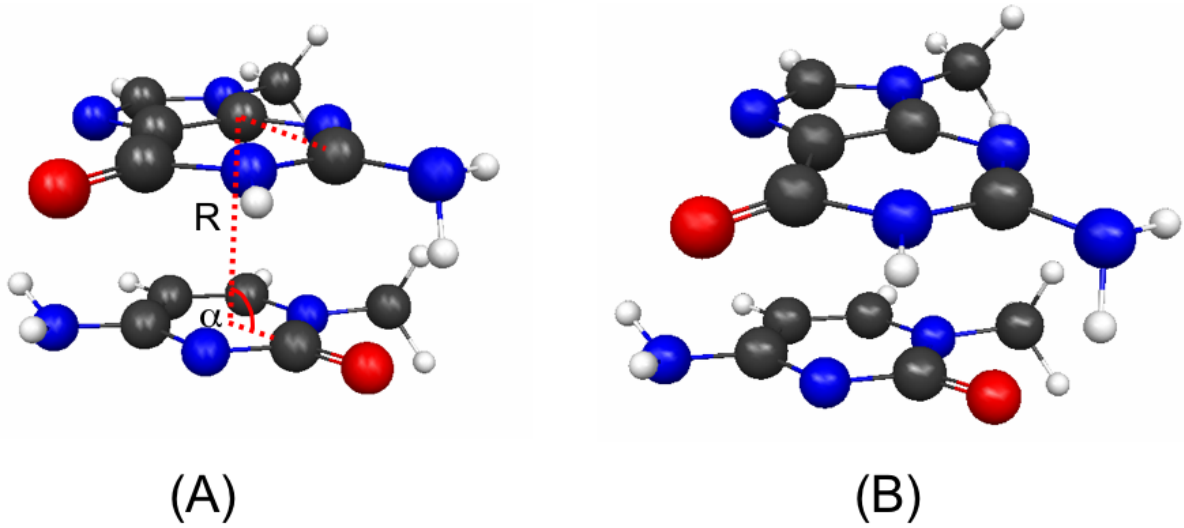


Fig. 5.6. Methylated guanine-cytosine stacked complex. (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref. 1. (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref. 1. Intermonomer distance R, angle α (chosen to demonstrate the relative displacement of the nucleotide bases with respect to one another), and the dihedral angle between the planes of the nucleobases (dashed red line) are given in Table 5.3 for each complex.

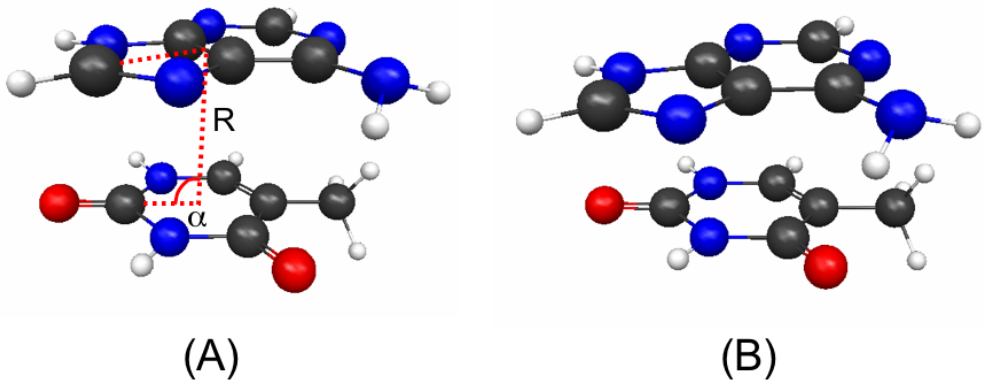


Fig. 5.7. Adenine-thymine stacked complex. (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref. 1. (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref. 1. Intermonomer distance R, angle α (chosen to demonstrate the relative displacement of the nucleotide bases with respect to one another), and the dihedral angle between the planes of the nucleotide bases (dashed red line) are given in Table 5.3 for each complex.

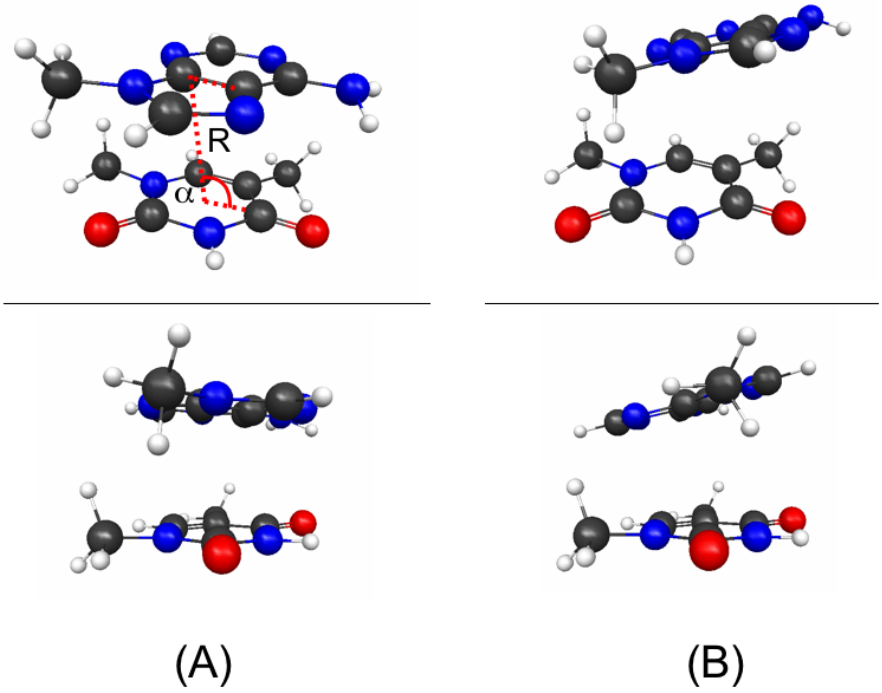


Fig. 5.8. Methylated adenine-thymine stacked complex. Two views are shown for both (A) and (B) to better illustrate the relative orientation of the nucleotide bases. (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref. 1. (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref. 1. Intermonomer distance R, angle α (chosen to demonstrate the relative displacement of the nucleotide bases with respect to one another), and the dihedral angle between the planes of the nucleotide bases (dashed red line) are given in Table 5.3 for each complex.

Table 5.3. Geometries of stacked complexes. Lengths (in Å) and angles (in degrees) of the hydrogen bonds in the Watson-Crick (WC) complexes of Figs. 5.5 - 5.8. "Opt type" refers to the level of theory used for the geometry optimization, either RI-MP2 [ref. 1] or EFP2. Angle α (defined in Figs. 5.5A - 5.8A) indicates the relative displacement of the nucleotide bases. The dihedral angle (also defined in Figs. 5.5A - 5.8A) goes between ring planes of the nucleotide bases and indicates how parallel these ring planes are.

	opt type	R	α	dihedral
G...C S	MP2	3.47	85.5	10.3
	EFP2	3.54	95.6	4.0
mG...mC S	MP2	3.21	91.6	-11.4
	EFP2	3.47	78.1	-15.2
A...T S	MP2	3.18	87.6	29.8
	EFP2	3.53	83.2	27.8
mA...mT S	MP2	3.14	100.6	-35.8
	EFP2	3.32	91.1	30.7

method overbinds the G…C S complexes more than any other stacked complex except EFP2-optimized mA…mT S. The EFP2 interaction energy of RI-MP2 optimized[1] G…C S is 1.6 kcal/mol more strongly bound than the estimated CCSD(T) energy[1] (Table 5.4).

In the EFP2-optimized methylated complex mG…mC S (Fig. 5.6B), the guanine nucleotide base is displaced over the cytosine relative to its orientation in RI-MP2 optimized[1] mG…mC S (Fig. 5.6A). The angle α, a measure of relative displacement of the nucleotide bases, is 91.6° in RI-MP2 optimized[1] mG…mC S and 78.1° in EFP2-optimized mG…mC S (Table 5.3). However, the nucleotide base ring planes are similarly oriented in these two structures, with dihedral angles differing by less than 4° (Table 5.3). While the EFP2 interaction energy of the RI-MP2 optimized[1] mG…mC WC complex is 2.3 kcal/mol lower in magnitude than the estimated CCSD(T)[1] energy, the EFP2 interaction energy of the EFP2-optimized mG…mC WC complex is just 0.3 kcal/mol higher than the estimated CCSD(T)[1] energy (Table 5.4).

The EFP2-optimized A…T S structure (Fig. 5.7B) is the most similar among the stacked complexes to its RI-MP2[1] optimized counterpart (Fig. 5.7A). The distance between nucleotide bases in EFP2-optimized A…T S is 0.35 Å greater than in RI-MP2 optimized[1] A…T S, though the relative displacement (given by α) and dihedral angle between nucleotide base planes differ by just 4.4° and 2° (Table 5.3). The EFP2 interaction energy of EFP2-optimized A…T S is only 0.3 kcal/mol larger in magnitude than the estimated

Table 5.4. Energies of stacked complexes. EFP2 energy components (Coulomb, exchange-repulsion, polarization, dispersion) and total interaction energy for the stacked nucleotide base complexes shown in Figs. 5.5 - 5.8. "Opt type" refers to the level of theory used for the geometry optimization, either RI-MP2 [ref. 1] or EFP2. The estimated CCSD(T)/CBS interaction energies[1] are shown for complexes with geometries taken from ref. 1. Energies in kcal/mol.

	opt type	EFP2 COULOMB	EFP2 EX-REP	EFP2 POL	EFP2 DISP	EFP2 TOTAL	CCSD(T)
G…C S	MP2	-20.6	20.6	-3.1	-17.5	-20.6	-19.0
	EFP2	-22.1	22.2	-3.7	-18.1	-21.7	
mG…mC S	MP2	-21.9	30.3	-3.1	-23.4	-18.1	-20.4
	EFP2	-19.7	22.3	-3.4	-20.0	-20.7	
A…T S	MP2	-8.6	16.1	-0.7	-17.6	-10.7	-12.3
	EFP2	-10.9	13.9	-0.6	-14.9	-12.6	
mA…mT S	MP2	-11.3	26.3	-1.3	-25.1	-11.4	-14.6
	EFP2	-15.9	18.3	-1.1	-18.2	-16.8	

CCSD(T)[1] interaction energy, while for A…T S optimized with RI-MP2[1], it is 1.6 kcal/mol lower.

Among the stacked structures, as with the hydrogen bonded structures, the greatest discrepancy between RI-MP2[1] and EFP2-optimized geometries occurs for the methylated adenine-thymine complex. Also as in the case of mA…mT WC compared to other hydrogen bonded structures, the difference between the estimated CCSD(T)[1] and EFP2 interaction energies of RI-MP2 optimized[1] mA…mT S is the greatest of all of the stacked structures. Two views of mA…mT S are shown in Fig. 5.8, where it is evident that the nucleotide bases in mA…mT S optimized with EFP2 (Fig. 5.8B) are rotated relative to their orientation in mA…mT S optimized with RI-MP2[1] (Fig. 5.8A). The dihedral angle between the nucleotide base planes differs greatly between the two structures; this angle is -35.8° in RI-MP2[1] optimized mA…mT S and is +30.7° in EFP2-optimized mA…mT S (Table 5.3). Compared to the estimated CCSD(T) interaction energies[1], EFP2 under-binds the RI-MP2 optimized[1] mA…mT S complex by 3.2 kcal/mol. The EFP2-optimized mA…mT S structure is 2.2 kcal/mol more strongly bound compared to the estimated CCSD(T)[1] value.

While the hydrogen bonded structures are more strongly bound than their respective stacked counterparts, the stacked guanine-cytosine complexes have a greater interaction energy than the hydrogen bonded adenine-thymine complexes. The EFP2 interaction energy for the RI-MP2 optimized[1] G…C S structure is 6.6 kcal/mol more strongly bound than that of

the RI-MP2 optimized[1] A…T WC structure (2.1 kcal/mol more strongly bound with estimated CCSD(T)[1]) and 4.5 kcal/mol more strongly bound than that of RI-MP2 optimized[1] mA…mT WC (0.8 kcal/mol with estimated CCSD(T)[1]) (Tables 5.2 and 5.4). The RI-MP2 optimized[1] stacked methylated structure mG…mC S has an EFP2 interaction energy 4.1 kcal/mol greater (in magnitude) than RI-MP2 optimized[1] A…T WC (3.5 kcal/mol with estimated CCSD(T)[1]) and 2.0 kcal/mol greater than RI-MP2 optimized[1] mA…mT WC (2.2 kcal/mol with estimated CCSD(T)[1]) (Tables 5.2 and 5.4).

Dispersion is the single greatest attractive contribution to the EFP2 total interaction energy for the stacked adenine-thymine pairs, but not for guanine-cytosine (Table 5.4). In mA…mT S optimized with RI-MP2[1], the dispersion energy is -25.1 kcal/mol, the highest for any structure examined. This is more than twice the binding contribution of the Coulomb term (-11.3 kcal/mol). The magnitude of the EFP2 dispersion energy for RI-MP2 optimized[1] A…T S exceeds the magnitude of the EFP2 exchange-repulsion by 1.5 kcal/mol (Table 5.4). In the guanine-cytosine pairs, the Coulomb term predominates, although the dispersion term is also large (within 4 kcal/mol of the Coulomb contribution for both EFP2-optimized and RI-MP2 optimized[1] G…C S structures). The exchange-repulsion term equals or approximately equals the magnitude of the Coulomb attraction in both G…C S structures, so dispersion remains extremely important for binding in stacked guanine-cytosine pairs. Among the methylated stacked pairs, the magnitude of the dispersion term is greater than the magnitude of the Coulomb term in all four structures examined. The exchange-repulsion term in all four methylated stacked pairs is greater in magnitude than either the Coulomb or dispersion terms. So, in fact, all of the interaction types make significant contributions to the binding of the stacked pairs, except for the polarization, which remains small in all species.

Conclusions

The EFP2 method predicts the structures of the hydrogen bonded and stacked guanine-cytosine, methylated guanine-cytosine, and adenine-thymine pairs in good agreement with RI-MP2[1]. Greater discrepancy is found between EFP2 and RI-MP2 optimized geometries of methylated adenine-thymine complexes.

The EFP2 total interaction energies are in fair to excellent agreement with the corresponding estimated CCSD(T)[1] results. For the RI-MP2 optimized[1] hydrogen bonded (Watson-Crick, or WC) structures, the root mean square deviation (RMSD) between the EFP2 and est. CCSD(T)[1] total interaction energies is 2.0 kcal/mol, with a maximum unsigned difference of 2.9 kcal/mol (corresponding to A…T WC). The RMSD between the EFP2 and est. CCSD(T)[1] total interaction energies when the hydrogen bonded structures are optimized with EFP2 (whereas the est. CCSD(T)[1] values correspond to the RI-MP2 optimized[1] structures) is 1.1 kcal/mol. The maximum unsigned difference, still corresponding to the A…T WC structure, is 1.6 kcal/mol. Among the stacked structures, the RMSD between EFP2 and est. CCSD(T)[1] energies for structures optimized with RI-MP2[1] is 2.3 kcal/mol. The maximum unsigned difference, corresponding to mA…mT S, is 3.2 kcal/mol. Comparing EFP2 interaction energies for the EFP2-optimized stacked structures with est. CCSD(T)[1] interaction energies for the RI-MP2 optimized[1] structures gives an RMSD of 1.8 kcal/mol and a maximum unsigned difference of 2.7 kcal/mol (for G…C S).

An accurate description of the dispersion energy is essential to determine the binding energies of the nucleotide base pairs, even for the hydrogen bonded structures. While the Coulomb interaction is the predominant attractive energy term for hydrogen bonded structures, its magnitude is only ~1-3 kcal/mol larger than the exchange-repulsion for EFP2-optimized A…T WC and EFP2-optimized mA…mT WC. The EFP2 exchange-repulsion term slightly (~1 kcal/mol) exceeds the magnitude of the EFP2 Coulomb term in the RI-MP2 optimized[1] A…T WC and mA…mT WC structures. In the guanine-cytosine and methylated guanine-cytosine hydrogen bonded structures, the magnitude of the Coulomb term is ~9 kcal/mol larger than the exchange-repulsion term among the EFP2-optimized structures; however, this accounts for only ~30% of the total interaction energy, the remainder being comprised of the dispersion and polarization energies.

Acknowledgements

This work was supported by a grant from the Air Force Office of Scientific Research.

References

1) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys* **2008**, *8*, 1985.
2) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989,** *157*, 479.
3) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevents, W. *J. Phys. Chem. A* **2001**, *105*, 293.
   Jensen, J. H.; Gordon, M. S. *Mol. Phys.* 89, 1313 (1996).
4) Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.
5) Gordon, M. S.; Slipchenko, L. V.; Li, H.; Jensen, J. H. *Annual Reports in Comp. Chem.* **2007**, *3*, 177.
6) Smith, T.; Slipchenko, L. V.; Gordon, M. S. *J. Phys. Chem. A* **2008**, *112*, 5286.
7) Smith, Q. A.; Slipchenko, L. V.; Gordon, M. S. *J. Phys. Chem. A* (in press)
8) Ghosh, D.; Kosenkov, D.; Vanovschi, V.; Williams, C. F.; Herbert, J. M.; Gordon, M. S.; Schmidt, M. W.; Slipchenko, L. V.; Krylov, A. I. *J. Phys. Chem. A* **2010**, *114*, 12739.
9) Langner, K. M.; Sokalski, W. A.; Leszczynski, J. *J. Chem. Phys.* **2007**, *127*, 111102.
10) Möller, C; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
11) Cybulski, S. M.; Lytle, M. L. *J. Chem. Phys*., **2007**, *127*, 141102.
12) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
13) Johnson, E. R.; Becke, A. D. *J. Chem. Phys*. **2006**, *124*, 174104.
14) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Am. Chem. Soc*. **2006**, *128*, 11730.
15) Sedlák, R.; Jurečka, P.; Hobza, P. *J. Chem. Phys*. **2007**, *127*, 075104.
16) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem*. **1993**, *14*, 1347.
17) Stone, A. J. *The Theory of Intermolecular Forces.* Oxford University Press: Oxford, **1996**.
18) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128.
19) Slipchenko, L. V.; Gordon, M. S. *Mol. Phys.* **2009**, *107*, 999.
20) Jenson, J. H.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.
21) Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *103*, 379.
   Amos, R. D.; Handy, N. C.; Knowles, P. J.; Rice, J. E.; Stone, A. J. *J. Phys. Chem.* **1985**, *89*, 2186.
22) Li, H.; Gordon, M. S.; Jensen, J. H. *J. Chem. Phys.* **2006**, *124*, 214108.
23) EFP2 geometry optimizations failed to converge on planar hydrogen bonded structures when the numerical DMA was used.
24) Boys, S. F.; Bernardi, F. *Mol. Phys.* **2002**, *100*, 65.
25) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437.
   Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.

CHAPTER 6: THE DISPERSION INTERACTION IN COMBINED AB INITIO-EFFECTIVE FRAGMENT POTENTIAL SYSTEMS

Introduction

The dispersion interaction is an attractive force between atoms and molecules that is caused by the interactions of induced multipoles. Dispersion arises from the correlated movement of electrons; an instantaneous multipole on one molecule may induce a multipole on another molecule. While dispersion is generally a weak intermolecular force, it is the sole attractive force between neutral atoms and molecules that lack permanent multipole moments. For example, dispersion is singularly responsible for the attraction between atoms of the noble gases, allowing for their condensation at low temperatures. Because dispersion depends on atomic or molecular polarizability, the dispersion energy is significant between atoms or molecules with large, diffuse electron clouds, such as π clouds.

The general formula for the dispersion interaction between two (quantum mechanical, QM) molecules comes from the second-order term in intermolecular perturbation theory. Based on this functional form, a formula for the dispersion interaction between two molecules modeled with the effective fragment potential method (EFP) was previously[1] derived and implemented in the General Atomic and Molecular Electronic Structure System (GAMESS)[2]. However, a formula for the dispersion energy in mixed systems – those in which one molecule is modeled with EFP and another with a full *ab initio* (AI) QM method (e.g. Hartree-Fock, MP2) – does not yet exist. This chapter presents efforts toward deriving and coding a formula for the EFP-AI dispersion interaction.

The first section shows the derivation of the dispersion energy in its most general form, from intermolecular perturbation theory. The next section presents a derivation of the EFP-EFP dispersion term and gives an overview of its implementation in GAMESS. The final section details progress toward an EFP-AI dispersion energy derivation and discusses results from the code that has been completed thus far.

Dispersion interaction from perturbation theory

The general formula for the dispersion interaction between two closed-shell nondegenerate ground state molecules can be derived with Rayleigh-Schrödinger perturbation theory, following the method of Stone[3]. The unperturbed Hamiltonian is given by the sum of the Hamiltonians for the individual molecules, A and B:

$$\hat{H}_0 = \hat{H}_A + \hat{H}_B \tag{1}$$

The perturbation operator embodies all electrostatic interactions between the molecules. It can be represented in numerous ways, the simplest of which is

$$\hat{V} = \sum_{i \in A} \sum_{j \in B} \frac{q_i q_j}{R_{ij}} \tag{2}$$

where q_i is the charge on particle i (electron or nucleus) on molecule A and R_{ij} is the distance between i and j. A more convenient representation is the multipolar expansion:

$$\hat{V} = T^{AB} q^A q^B + \sum_{a}^{x,y,z} T_a^{AB} \left(q^A \mu_a^B - \mu_a^A q^B \right) - \sum_{a,b}^{x,y,z} T_{ab}^{AB} \mu_a^A \mu_b^B + \ldots \tag{3}$$

where q^A is the net charge on molecule A, μ_a^B is the ath directional (x, y, z) component of the dipole moment on B, etc. The quantities T, T_a, T_{ab}, etc. are the field, field gradient, field second derivative, etc.; they are electrostatic tensors of rank indicated by the number of subscripts. T corresponds to charge-charge interactions, T_a to charge-dipole interactions, T_{ab} to dipole-dipole, T_{abc} to dipole-quadrupole, and so on. The formulae of the first three tensors appear below, where R_{ij}^a denotes the ath directional component of the distance between i and j.

$$T^{ij} = \frac{1}{R_{ij}} \tag{5}$$

$$T_a^{ij} = \nabla_a \frac{1}{R} = -\frac{\vec{R}_{ij}^a}{R_{ij}^3} \tag{6}$$

$$T_{ab}^{ij} = \nabla_a \nabla_b \frac{1}{R} = \frac{3\vec{R}_{ij}^a \vec{R}_{ij}^b - R_{ij}^2 \delta_{ab}}{R_{ij}^5} \tag{7}$$

When the perturbation expansion is carried out, the zeroth-order term gives the sum of the ground state energies of A and B, and the first-order term gives the Coulomb

(electrostatic) interaction energy. The second-order energy term in the perturbation expansion is given by

$$W_0'' = -\sum_{m,n} \frac{\langle 0_A 0_B | \hat{V} | mn \rangle \langle mn | \hat{V} | 0_A 0_B \rangle}{W_{mn} - W_{0_A 0_B}} \tag{8}$$

where m and n are states of molecules A and B, respectively, 0_A represents the ground state of molecule A, and $W_{mn} = E_m + E_n$ is the energy of the system in state $|mn\rangle$. In Eq. (8), either m or n may be zero, but not both. This second-order perturbative term encompasses both the induction (polarization) and dispersion energies. Induction arises from the interactions between permanent multipoles on one molecule and induced multipoles on the other. As such, its representation in Eq. (8) occurs in those terms in the summation in which either m or n is equal to 0, i.e., one of the interacting molecules is in its ground state, while the other is in an (induced) excited state. The remainder of the terms in the summation, those in which both m and n refer to excited states, correspond to the dispersion energy:

$$E^{disp} = -\sum_{\substack{m \neq 0 \\ n \neq 0}} \frac{\langle 0_A 0_B | \hat{V} | mn \rangle \langle mn | \hat{V} | 0_A 0_B \rangle}{E_m^A + E_n^B - E_0^A - E_0^B} \tag{9}$$

The multipolar expansion of Eq. (3) may then be substituted for the perturbation operator in Eq. (9). For simplicity, only the dipole-dipole term of the expansion is shown here. Since the charges q in Eq. (3) are scalar values, integrals involving q will be of the form, for example, $\langle 0_A | q^A | m \rangle = q^A \langle 0_A | m \rangle$; these integrals are zero because the ground and excited states are orthogonal to each other. Therefore, the multipole expansion, when used in this context, properly begins with the dipole-dipole term. The dispersion energy corresponding to the dipole-dipole interaction only is labeled E_6^{disp} for reasons that will be shown at the end of the derivation.

$$E_6^{disp} = -\sum_{\substack{m \neq 0 \\ n \neq 0}} \frac{\langle 0_A 0_B | \sum_{a,b}^{x,y,z} T_{ab}^{AB} \hat{\mu}_a^A \hat{\mu}_b^B | mn \rangle \langle mn | \sum_{c,d}^{x,y,z} T_{cd}^{AB} \hat{\mu}_c^A \hat{\mu}_d^B | 0_A 0_B \rangle}{E_m^A + E_n^B - E_0^A - E_0^B} \tag{10}$$

The dipole operators apply only to the states of their respective molecules (m on A or n on B), and the quadrupole tensors do not depend on the states. Eq. (10) may then be simplified to

$$E_6^{disp} = -\sum_{abcd}^{x,y,z} T_{ab}^{AB} T_{cd}^{AB} \sum_{\substack{m\neq 0 \\ n\neq 0}} \frac{1}{E_{m0}^{A} + E_{n0}^{B}} \langle 0_A | \hat{\mu}_a^A | m \rangle \langle m | \hat{\mu}_c^A | 0_A \rangle \langle 0_B | \hat{\mu}_b^B | n \rangle \langle n | \hat{\mu}_d^B | 0_B \rangle \tag{11}$$

where $E_{m0} = E_m - E_0$. These factors do not permit the easy separation of Eq. (11) into portions relating to molecule *A* and portions relating to molecule *B*. The identity[3]

$$\frac{1}{A+B} = \frac{2}{\pi} \int_0^\infty \frac{AB}{\left(A^2 + \omega^2\right)\left(B^2 + \omega^2\right)} d\omega \tag{12}$$

may be applied to the denominator of Eq. (11) to obtain

$$E_6^{disp} = -\frac{2\hbar}{\pi} \sum_{abcd}^{x,y,z} T_{ab}^{AB} T_{cd}^{AB} \int_0^\infty d\omega \sum_{\substack{m\neq 0 \\ n\neq 0}} \frac{\omega_m^A \langle 0_A | \hat{\mu}_a^A | m \rangle \langle m | \hat{\mu}_c^A | 0_A \rangle}{\hbar\left(\left(\omega_m^A\right)^2 + \omega^2\right)} \frac{\omega_n^B \langle 0_B | \hat{\mu}_b^B | n \rangle \langle n | \hat{\mu}_d^B | 0_B \rangle}{\hbar\left(\left(\omega_n^B\right)^2 + \omega^2\right)} \tag{13}$$

after the energies in Eq. (11) are expressed in terms of frequencies, e.g. $E_{m0}^A = \hbar\omega_m^A$. From time-dependent perturbation theory, it can be shown that the response of a molecule to an oscillating electric field is an oscillating dipole moment. If a field F_b with frequency ω is given by $F_b e^{-i\omega t}$, then the dipole moment is $\mu_a = \alpha_{ab}(\omega) F_b e^{-i\omega t}$. The components of the frequency-dependent dynamic polarizability tensor, α, are given by[3]

$$\alpha_{ab}(\omega) = \sum_m \frac{\omega_m \left\{ \langle 0 | \hat{\mu}_a | m \rangle \langle m | \hat{\mu}_b | 0 \rangle + \langle 0 | \hat{\mu}_b | m \rangle \langle m | \hat{\mu}_a | 0 \rangle \right\}}{\hbar\left(\omega_m^2 - \omega^2\right)} = 2\sum_m \frac{\omega_m \langle 0 | \hat{\mu}_a | m \rangle \langle m | \hat{\mu}_b | 0 \rangle}{\hbar\left(\omega_m^2 - \omega^2\right)} \tag{14}$$

α describes the propagation of a density fluctuation within a molecule.[4] Expressing the factor ω^2 as $\omega^2 = -\left(-\omega^2\right) = -(i\omega)^2$, Eq. (13) can be recast in terms of the dynamic polarizability tensor at an imaginary frequency:

$$E_6^{disp} = -\frac{\hbar}{2\pi} \sum_{abcd}^{x,y,z} T_{ab}^{AB} T_{cd}^{AB} \int_0^\infty \alpha_{ac}^A(i\omega) \alpha_{bd}^B(i\omega) d\omega \tag{15}$$

The concept of imaginary frequencies is purely a mathematical construct, with no real physical meaning. Regardless, the integral in Eq. (15) is well-behaved and decreases monotonically to 0 as $\omega \to \infty$.

The dynamic polarizability tensors in Eq. (15) are implicitly calculated at a single point on each molecule. However, there are known issues with convergence when using this approach; to be effective, it may require higher terms in the multipolar expansion[5] (i.e. quadrupole polarizabilities might need to be considered). A distributed polarizability model,

in which polarizability tensors are calculated at multiple expansion points, has the advantage of more successful convergence and possibly giving a more realistic portrayal of the response of a molecule to the nonuniform fields arising from the interacting molecules[5]. Converting Eq. (15) to the distributed polarizability model gives

$$E_6^{disp} = -\frac{\hbar}{2\pi}\sum_{i\in A}\sum_{j\in B}\sum_{abcd}^{x,y,z} T_{ab}^{ij} T_{cd}^{ij} \int_0^{\infty} \alpha_{ac}^{i}(i\omega)\alpha_{bd}^{j}(i\omega)d\omega \tag{16}$$

Eq. (16) is the final expression for the dipole-dipole term of the dispersion interaction energy, where *i* and *j* are expansion points for distributed polarizabilities on molecules *A* and *B*.

The preceding derivation relied only on the dipole-dipole term in the multipolar expansion of the interaction operator; the higher-order terms may be derived similarly. The dipole-dipole interaction tensors T_{ab} in Eq. (16) each scale as R^{-3}, so the distance dependence of E_6^{disp} scales as R^{-6}. The next term in the dispersion energy expansion arises from dipole-dipole interactions T_{ab} on one molecule and dipole-quadrupole interactions T_{cde} on the other. The latter tensor scales as R^{-4}. This implies that the distance dependence of the corresponding dispersion energy term scales as R^{-7}. The scaling associated with dipole-quadrupole interactions on both molecules is then R^{-8}, and so on. Thus, the total dispersion energy can be expressed in the form

$$E^{disp} = \sum_{n\geq 6}\frac{C_n}{R^n} \tag{17}$$

However, the odd terms are almost always ignored. For atoms and for molecules with an inversion center, the dipole-quadrupole electrostatic tensor components are zero, as are the components of all higher-order tensors of paired "unlike" multipoles. While, in actuality, the odd terms in Eq. (17) do contribute to the total dispersion energy of interacting nonsymmetric molecules, the contribution is assumed to be insignificant. Therefore, the usual form of the dispersion expansion is

$$E^{disp} = \frac{C_6}{R^6} + \frac{C_8}{R^8} + \frac{C_{10}}{R^{10}} + \ldots \tag{18}$$

Dispersion interaction between effective fragment potentials

The equation for the dispersion interaction between two molecules modeled with the effective fragment potential method begins with Eq. (18). The dispersion energy expression is truncated after the first (dipole-dipole) term; the remainder of the series has been shown[1] to be well approximated as $^1/_3$ of the value of the first term. The distributed polarizability expression of Eq. (16) is the foundation for calculating the dipole-dipole term. EFP expansion points for the dispersion energy are taken to be the localized molecular orbital (LMO) centroids, i.e. the centers of charge, for bonds and lone pairs. LMO centroids are a physically meaningful choice for distributed polarizability points because the induced dipole moment is related to a shift in the position of the centroids. The distance factor R^{-6} is calculated between these LMO centroids, leaving the C_6 coefficient (the remaining portion of Eq. (16)) to be determined:

$$C_6 = \sum_{i\in A}\sum_{j\in B}\sum_{abcd}^{x,y,z}\int_0^\infty \alpha_{ac}^{i}(i\omega)\alpha_{bd}^{j}(i\omega)d\omega \tag{19}$$

The superscripts *i* and *j* refer to LMOs on fragments *A* and *B*, respectively.

Following the approach of Amos[6,7], the dynamic polarizability tensors α associated with each polarizable point are computed via the time-dependent analog of the coupled perturbed Hartree-Fock (CPHF)[8] equations. Details may be found in refs. 1 and 6 - 8. Specifically, each α tensor is computed at 12 distinct imaginary frequency values for later use with integration using numerical quadrature.

The C_6 coefficient given in Eq. (19) is anisotropic, as the summation is over all possible x, y, z pairs. In most computational methods, an isotropic C_6 – one which does not include xy, xz, and other off-diagonal terms – is preferred. This is because only the trace of the polarizability tensor is distance-independent, so it remains unchanged during geometry optimizations; otherwise, the calculation of the tensors would have to be repeated at each optimization step, a process that is too time-consuming. Therefore, an isotropic approximation is employed, in which the summation over distance components is eliminated by expressing α for each LMO as $^1/_3$ of the trace of the polarizability tensor at a given

imaginary frequency. This value is represented by $\overline{\alpha}^{j}(i\omega)$ for LMO j and frequency ω. The isotropic C_6 becomes

$$C_6 = \sum_{i\in A}\sum_{j\in B}\int_0^\infty \overline{\alpha}^{i}(i\omega)\overline{\alpha}^{j}(i\omega)d\omega \tag{20}$$

To convert the integral to a range conducive with the use of Gauss-Legendre numerical quadrature, a substitution of variables is made:

$$\omega = \nu_0\frac{(1-t)}{(1+t)} \qquad\qquad d\omega = \frac{2\nu_0 dt}{(1+t)^2} \tag{21}$$

This converts the range of integration to be from -1 to +1. By using Gauss-Legendre abscissas for t, it also determines the values of the (imaginary) frequencies at which the polarizability tensors must be calculated. This substitution, including the optimal value of ν_0 (0.3), was determined in ref. 9.

The numerical quadrature scheme for the C_6 coefficient between LMOs i and j is given by

$$C_6^{ij} = \sum_{k=1}^{12} w_k \frac{2\nu_0}{(1-t_k)^2}\overline{\alpha}^{i}(i\omega_k)\overline{\alpha}^{j}(i\omega_k) \tag{22}$$

where t_k and w_k are Gauss-Legendre abscissas and weights and $\overline{\alpha}^{j}(i\omega_k)$ is $^1/_3$ of the trace of the polarizability tensor for LMO j at frequency ω_k (given by Eq. (21) for each t_k).

Once the coefficients have been determined, the dispersion energy has the form

$$E_{EFP-EFP}^{disp} = -\frac{4}{3}\sum_{i\in A}\sum_{j\in B}\frac{C_6^{ij}}{R_{ij}^6} \tag{23}$$

where the factor of $^4/_3$ comes from the approximation of the higher-order dispersion terms as $^1/_3$ the value of the dipole-dipole term[9].

In the implementation of EFP-EFP dispersion coded in GAMESS, the dispersion energy is multiplied by a distance-dependent damping function in order to account for charge penetration effects that occur at short inter-fragment separations. Two damping functions are available in the GAMESS EFP code. The original implementation of the EFP-EFP dispersion interaction utilized a Tang-Toennies[10] damping function

$$f_6(R) = 1 - e^{-\beta R}\sum_{k=0}^{6}\frac{(\beta R)^k}{k!} \tag{24}$$

where β is a parameter assigned the value of 1.5 [ref. 9]. A newer damping function[11], free of parameters and depending explicitly on the overlap matrix *S* between fragment charge clouds, is given by

$$C_6^{ij} \rightarrow \left(1 - \left|S^{ij}\right|^2 \left(1 - 2\ln\left|S^{ij}\right| + 2\ln^2\left|S^{ij}\right|\right)\right) C_6^{ij} \tag{25}$$

Dispersion interaction between an effective fragment potential and an *ab initio* molecule

In constructing an expression for the EFP-AI dispersion interaction, it would be appealing to obtain an equation analogous with Eq. (16),

$$E_6^{disp} = -\frac{\hbar}{2\pi} \sum_{i \in A} \sum_{j \in B} \sum_{abcd}^{x,y,z} T_{ab}^{ij} T_{cd}^{ij} \int_0^{\infty} \alpha_{ac}^{i}(i\omega) \alpha_{bd}^{j}(i\omega) d\omega$$

as was done with the EFP-EFP dispersion derivation. In the case of EFP-EFP, the T_{ab} and T_{cd} tensors become a straightforward R^{-6} calculation between fragment LMO centroids and the remainder of the expression (the integral over the imaginary frequencies) becomes a C_6^{ij} value calculated between each LMO pair. The calculation of the dynamic polarizability tensors α at the necessary (pre-determined) frequencies is part of the MAKEFP procedure, which generates the fragment potential parameters before the calculation of interaction energy components begins. However, this procedure is too time-consuming to be appropriate for the AI molecule, since constructing the α tensors would necessarily occur during the calculation of the EFP-AI interaction. For example, if a Hartree-Fock calculation on phenol with the 6-311++G(3df,2p) basis set (equivalent to 375 basis functions) takes approximately six minutes on a given computer, a CPHF calculation on the same chemical system requires 10 minutes, and a time-dependent Hartree-Fock (TDHF) calculation requires 12 minutes. To obtain the polarizability tensors on the AI molecule, both CPHF and TDHF calculations would need to be performed, thus increasing the run time significantly. Therefore, the EFP-AI dispersion derivation must diverge from the perturbation theory derivation before both of the α tensors are constructed.

Let *A* refer to the AI molecule and *B* to the EFP fragment. The derivation of the EFP-AI dispersion interaction begins with Eq. (13) in the perturbation theory derivation of the dispersion energy:

$$E_6^{disp} = -\frac{2\hbar}{\pi}\sum_{abcd}^{x,y,z} T_{ab}^{AB} T_{cd}^{AB} \int_0^\infty d\omega \sum_{\substack{m\neq 0 \\ n\neq 0}} \frac{\omega_m^A \langle 0_A | \hat{\mu}_a^A | m \rangle \langle m | \hat{\mu}_c^A | 0_A \rangle}{\hbar\left(\left(\omega_m^A\right)^2 + \omega^2\right)} \frac{\omega_n^B \langle 0_B | \hat{\mu}_b^B | n \rangle \langle n | \hat{\mu}_d^B | 0_B \rangle}{\hbar\left(\left(\omega_n^B\right)^2 + \omega^2\right)}$$

As before, *m* and *n* are excited states of molecules *A* and *B*, T_{ab} is the electrostatic tensor given by Eq. (7), $\hat{\mu}_a^A$ is the dipole operator pertaining to molecule *A*, and frequency ω_m^A for molecule *A* is given by the relation $E_{m0_A}^A = \hbar\omega_m^A$ where $E_{m0_A}^A$ is the difference in energy between the ground state 0_A and the excited state *m*. Since values of the dynamic polarizability tensors are calculated for EFP fragments as part of the MAKEFP process, the EFP fragment *B* may be treated in much the same way as it is in modeling the EFP-EFP dispersion interaction. Therefore, using the definition of the α tensor given in Eq. (14), the frequencies and dipole integrals pertaining to EFP fragment *B* are expressed as a polarizability tensor:

$$E_6^{EFP-AI} = -\frac{\hbar}{\pi}\sum_{abcd}^{x,y,z} T_{ab}^{AB} T_{cd}^{AB} \int_0^\infty d\omega \sum_{m\neq 0} \frac{\omega_m^A \langle 0_A | \hat{\mu}_a^A | m \rangle \langle m | \hat{\mu}_c^A | 0_A \rangle}{\hbar\left(\left(\omega_m^A\right)^2 + \omega^2\right)} \alpha_{bd}^B(i\omega)$$

$$= -\frac{1}{\pi}\sum_{abcd}^{x,y,z} T_{ab}^{AB} T_{cd}^{AB} \sum_{m\neq 0} \langle 0_A | \hat{\mu}_a^A | m \rangle \langle m | \hat{\mu}_c^A | 0_A \rangle \int_0^\infty d\omega \frac{\omega_m^A}{\left(\omega_m^A\right)^2 + \omega^2} \alpha_{bd}^B(i\omega) \qquad (26)$$

Three considerations must be taken into account in order to convert Eq. (26) into a computationally useful form. First, the equation must be re-cast in terms of orbitals rather than the sum-over-states form. Second, distributed polarizabilities must be used with the EFP fragment, *B*. Finally, the isotropic approximation must be employed.

To convert from a sum-over-states approach to an orbital-based approach, let *i* correspond to occupied molecular orbitals (MOs) and *k* to virtual MOs on the AI molecule. Also let $\omega_{ki} = \omega_k - \omega_i$. This changes the dipole integral and the integral over imaginary frequencies in Eq. (26) to:

$$\sum_{m\neq 0} \langle 0_A | \hat{\mu}_a^A | m \rangle \langle m | \hat{\mu}_c^A | 0_A \rangle \rightarrow \sum_i^{occ} \sum_k^{vir} \langle i | \hat{\mu}_a | k \rangle \langle k | \hat{\mu}_c | i \rangle \qquad (27)$$

$$\int_0^\infty d\omega \frac{\omega_m^A}{\left(\omega_m^A\right)^2+\omega^2}\alpha_{bd}^B(i\omega) \rightarrow \int_0^\infty d\omega \frac{\omega_{ki}^A}{\left(\omega_{ki}^A\right)^2+\omega^2}\alpha_{bd}^B(i\omega) \tag{28}$$

The use of distributed polarizabilities is desired on the EFP part, with expansion points at LMO centroids *j* on molecule *B*. This entails the conversion

$$\alpha_{bd}^B(i\omega) \rightarrow \sum_{j\in B}\alpha_{bd}^j(i\omega) \tag{29}$$

Changing from sum-over-states to the orbital approximation and changing to distributed multipoles on fragment *B*, Eq. (26) becomes

$$E_6^{EFP-AI} = -\frac{1}{\pi}\sum_{j\in B}\sum_{abcd}^{x,y,z}\sum_i^{occ}\sum_k^{vir} T_{ab}^{ij}T_{cd}^{ij}\langle i|\hat{\mu}_a|k\rangle\langle k|\hat{\mu}_c|i\rangle\int_0^\infty d\omega \frac{\omega_{ki}^A}{\left(\omega_{ki}^A\right)^2+\omega^2}\alpha_{bd}^j(i\omega) \tag{30}$$

The isotropic approximation is then employed to eliminate off-diagonal terms in the tensors:

$$E_6^{EFP-AI} = -\frac{1}{\pi}\sum_{j\in B}\sum_{a}^{x,y,z}\sum_i^{occ}\sum_k^{vir} T_{aa}^{ij}T_{aa}^{ij}\langle i|\hat{\mu}_a|k\rangle\langle k|\hat{\mu}_a|i\rangle\int_0^\infty d\omega \frac{\omega_{ki}^A}{\left(\omega_{ki}^A\right)^2+\omega^2}\overline{\alpha}^j(i\omega) \tag{31}$$

where $\overline{\alpha}^j(i\omega)$ is $^1/_3$ of the trace of the polarizability tensor for LMO *j* and frequency ω. Now, $T_{aa}^{ij}T_{aa}^{ij} = \left(T_{aa}^{ij}\right)^2 = \left(\frac{1}{R_{ij}^3}\right)^2 = \frac{1}{R_{ij}^6}$, where R_{ij} is the distance between centroids of LMOs *i* and *j*, and the remainder of the terms inside the summations in Eq. (31) corresponds to the distributed C_6 coefficient:

$$C_6^{ij} = \langle i|\hat{\mu}_a|k\rangle\langle k|\hat{\mu}_a|i\rangle\int_0^\infty d\omega \frac{\omega_{ki}^A}{\left(\omega_{ki}^A\right)^2+\omega^2}\overline{\alpha}^j(i\omega) \tag{32}$$

Implementation

Eq. (31) is a MO-based expression for the EFP-AI dispersion interaction; however, it would be advantageous to express the dipole integral in terms of atomic orbitals (AOs). The dipole integrals in the AO basis are readily available (calculated upon Hartree-Fock convergence). Therefore, the dipole integrals are computed in terms of AOs and converted to the MO basis for use with the remainder of the calculation. Only the occupied × virtual block

of the matrix of dipole integrals is required, so the basis conversion takes the form of the transformation

$$\mu_a^{A,occ\times vir} = OCC^T \cdot \mu_a^A \cdot VIR \tag{33}$$

where μ_a^A is the complete matrix of dipole integrals with elements $\langle a|\hat{\mu}_a^A|b\rangle$ over all AOs *a* and *b*, *OCC* consists of the first N_A (= number of occupied orbitals) columns of the MO coefficient matrix, and *VIR* consists of the remaining N_{total} - N_A columns. This produces a matrix $\mu_a^{A,occ\times vir}$ with dimensions $N_A \times (N_{total} - N_A)$ corresponding to the dipole integrals between occupied × virtual canonical MOs. Rather than using all of the occupied MOs in the calculation of the distributed C_6 values, only the valence LMOs are used.

Code for the integral over imaginary frequencies was obtained through straightforward modification of the EFP-EFP C_6 code, used to calculate Eq. (20). The same transformation of variables (Eq. (21)) was used with Gauss-Legendre numerical integration, resulting in a formula similar to Eq. (22). Unlike Eq. (22), the α polarizability tensor on molecule *A* is replaced with an expression dependent on MO energy differences. (Since the MO energies and the factors ω_{ki} in Eq. (31) differ by $\hbar$, which is 1 in atomic units, the energy difference between virtual MO *k* and occupied MO *i* is used for ω_{ki}.) The EFP-AI expression for the imaginary frequency integral is

$$I^j(E_{ki}) = \sum_{k=1}^{12} w_k \frac{2\nu_0}{(1-t_k)^2} \cdot \frac{E_{ki}}{E_{ki}^2 + \omega_k^2} \overline{\alpha}^j(i\omega_k) \quad \text{where} \quad \omega_k = \nu_0 \frac{(1-t_k)}{(1+t_k)} \tag{34}$$

As shown in Eq. (32), the expression in Eq. (34) must be multiplied by $\langle i|\hat{\mu}_a|k\rangle\langle k|\hat{\mu}_a|i\rangle$ to produce the distributed C_6 coefficients. First, for each *i* and *k*, I^j of Eq. (34) is multiplied by $\langle i|\hat{\mu}_a|k\rangle$ (terms in the occupied × virtual canonical MO-basis dipole integral matrix). The resulting quantity is multiplied by the transpose of the occupied × virtual MO-basis dipole integral matrix, using matrix multiplication (dot product). Element (i,i') of the resulting matrix V_a^j is given by

$$\left(V_a^j\right)_{i,i'} = \sum_k^{vir} \langle i|\mu_a|k\rangle \left[\int_0^\infty d\omega \frac{E_{ki}}{E_{ki}^2 + \omega^2} \overline{\alpha}^j(i\omega)\right] \langle k|\mu_a|i'\rangle \tag{35}$$

Since the $\langle i|\hat{\mu}_a|k\rangle$ elements obtained thus far are integrals between canonical MOs, a transformation must be applied to V_a^j, in the form of multiplication by a localization matrix,

to give distributed C_6 coefficients corresponding to LMOs. Finally, these C_6 coefficients can be multiplied by R^{-6} calculated between the centroids of LMO *i* on *A* and LMO *j* on *B*.

Results and Discussion

The neon dimer provides a concrete example of the process for determining C_6 coefficients in EFP-AI systems. Having four valence orbitals (lone pairs), a neon atom is modeled with EFP as four α polarizability tensors on the orbital centroids. The EFP-EFP treatment of dispersion entails integrals between each α on one fragment with every α on the other fragment, giving 16 C_6 integrals. The EFP-AI approach outlined above produces four matrices (one for each α^j on the EFP part) of dimension 4 × 4 (valence orbitals). The diagonal elements of these four matrices give the effective C_6 values between the two valence orbitals on the AI molecule and the EFP expansion point. For comparison, the distributed EFP-EFP C_6 coefficients for the neon dimer, calculated with basis set 6-311+G(2p), are all approximately 0.01931, differing from one another after the 8th decimal place. The effective EFP-AI coefficients, with both the AI and the EFP part calculated in the 6-311+G(2p) basis set, are all approximately 0.01586, differing from each other after the 6th decimal place.

Upon transformation to the LMO basis, the diagonal elements of the localized matrices correspond to isotropic EFP-EFP C_6^{ij} values for various dimer systems (Table 6.1).

Table 6.1. Comparison of EFP-EFP and EFP-AI distributed C_6 dispersion coefficients.

Dimer type	EFP-EFP C_6	EFP-AI C_6
Ne		
6-311+G(2p)	0.01931	0.01586
Ar		
6-31+G(2d,2p)	2.477	2.114
6-311+G(3d,3p)	3.762	3.141
CH_4		
STO-3G	1.36	1.26

While the calculation of R^{-6} between LMOs has not yet been implemented, the similarity between the EFP-EFP C_6 and the distributed C_6 values of EFP-AI suggests that the method presented here is correct.

Conclusions

EFP-AI distributed C_6^{ij} coefficients can be obtained by transforming AI dipole integrals from the atomic orbital basis to the localized molecular orbital (LMO) basis, then multiplying by an integral between the EFP dynamic polarizability tensor and a function of AI orbital energies. This gives a C_6 value for each pair of LMOs *i* and *j*, on the *ab initio* molecule and the EFP fragment, respectively. For all systems examined, this method produces values similar to the distributed C_6 coefficients in the EFP-EFP dispersion interaction. Total EFP-AI dispersion interaction energies can be obtained by calculating values of R^{-6} where R is the distance between LMOs *i* and *j*.

References

1) Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *103*, 379.
2) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem*. **1993**, *14*, 1347.
3) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: New York/Oxford, **1996**.
4) McWeeny, R. *Methods of Molecular Quantum Mechanics*, 2nd ed; Academic Press: London, **1989**.
5) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.
6) Amos, R. D.; Handy, N. C.; Knowles, P. J.; Rice, J. E.; Stone, A. J. *J. Phys. Chem.* **1985**, *89*, 2186.
7) Ioannou, A. G.; Colwell, S. M.; Amos, R. D. *Chem. Phys. Lett.*, **1997**, *278*, 278.
8) Yamaguchi, Y.; Goddard, J. D.; Osamura, Y.; Schaefer, H. F. *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*; Oxford University Press: New York/Oxford, **1994**.
9) Gross. E. K. U.; Ullrich, C. A.; Gossmann, U. J. *NATO ASI Series, Series B: Phys.* **1995**, *337*, 149.
10) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726.

11) Slipchenko, L. V.; Gordon, M. S. *Mol. Phys.* **2009**, *107*, 999.
12) Glaesemann, K. R.; Gordon, M. S. *J. Chem. Phys.* **1998**, *108*, 9959.
Glaesemann, K. R.; Gordon, M. S. *J. Chem. Phys.* **1999**, *110*, 6580.

CHAPTER 7: CONCLUSIONS

The geometries of small aluminum clusters Al_{13} and Al_{13}^- were examined in Chapter 2 with both *ab initio* and density functional (DFT) methods and with various basis sets. Al_{13}^- is a closed-shell species, while Al_{13} is a doublet radical. Second-order perturbation theory (MP2) with the largest basis set used, 6-311+G(2d), finds a distorted icosahedral ground state structure for Al_{13} and a perfect icosahedral structure for the anion. CCSD(T) energies were computed for the neutral and anionic clusters at their respective MP2 optimized geometries. The difference in the energies gives the adiabatic electron affinity of Al_{13}. Using the CCSD(T) energies, this value was determined to be 3.57 eV, which compares favorably with the experimentally determined value of 3.62 ± 0.06 eV. The close agreement implies that the MP2 geometries are correct. The MP2 geometries were compared with geometries obtained from DFT calculations. Different combinations of functionals and basis sets were found to produce different results with DFT. While the BPW91 functional predicts ground state icosahedral structures, like those found with MP2, the B3LYP functional finds icosahedral and decahedral structures that are nearly isoenergetic. When using the LANL2DZ functional with a small basis set, decahedral structures are favored. An MP2 Hessian (second derivative of the potential energy with respect to the coordinates) calculation demonstrated that the decahedral structure is a transition state.

The general effective fragment potential method (EFP2) was used to model substituted benzene dimers in Chapter 3 and pyridine and benzene-pyridine dimers in Chapter 4. In both studies, EFP2 was first used to model a series of constrained (non-equilibrium) structures that were specifically chosen to examine π-π interactions. These structures included T-shaped, sandwich, and parallel displaced geometries. Because the structures had previously been examined with MP2, SCS-MP2, CCSD(T), and/or SAPT, high-level *ab initio* data was available for comparison. EFP2 energies for the constrained structures compare very well with the *ab initio* energies, frequently differing by less than a few tenths of a kcal/mol. Compared to SAPT, EFP2 consistently – but slightly – overestimates the magnitude of the dispersion energy term. For the majority of structures, EFP2 favors a slightly larger intermonomer separation than MP2 or CCSD(T) does; the EFP2

underestimation of dimer binding energy increases in degree as intermonomer separation grows smaller. Overall, EFP2 was shown to be in excellent agreement with high-level *ab initio* methods for modeling these dispersion-dominated systems. The agreement is especially impressive when taking into account the miniscule run-time of an EFP2 calculation (less than one second, once the fragment parameters have been generated beforehand).

In addition to comparing EFP2 results with high-level *ab initio* data, Chapters 3 and 4 also showcase the utility of EFP2, in combination with Monte Carlo/simulated annealing (MC/SA), for exploring the potential energy surfaces of the dimers. Three major structural motifs were found in common with all of the substituted benzene and pyridine dimers: offset T-shaped, parallel displaced, and edge-on. The offset T-shaped configuration involves a "π-hydrogen" interaction between a hydrogen on one monomer and the ring center (π cloud) of the other. The parallel displaced configuration is generally similar to the constrained parallel displaced structure examined in the first part of each study, but the monomers are not usually perfectly parallel. Additionally, in the case of the substituted benzene dimers, the substituent lies across the ring of the benzene monomer; this type of configuration was not among those examined in the first part of the study. The relative energies of these structural motifs vary. For benzene-pyridine and pyridine dimers, the edge-on motif describes the global minimum. For benzene-benzonitrile and benzene-fluorobenzene, the global minimum is parallel displaced. For benzene-phenol and the unsubstituted benzene dimer, the T-shaped structure is the most energetically preferred. The edge-on configuration usually involves a larger attractive Coulomb (electrostatic) term than occurs in the other configurations, indicating an interaction between the partially negative substituent or heteroatom and the partially positive hydrogens of the second monomer. Although the Coulomb term is usually the largest attractive component of the interaction energy for the edge-on structures, the dispersion term is significant as well. In the unsubstituted benzene dimer, this structure is stabilized primarily by dispersion energy.

For both hydrogen bonded and stacked motifs of guanine-cytosine, methylated guanine-cytosine, and adenine-thymine nucleotide base pairs, EFP2 predicts structures very similar to those found with MP2 optimizations. A greater difference between EFP2 and MP2 optimized structures occurs with methylated adenine-thymine pairs. EFP2 total interaction

energies for these structures are in good agreement with estimated CCSD(T) interaction energies. The root mean square deviation (RMSD) between EFP2 and est. CCSD(T) single point energies for hydrogen bonded structures held at their MP2 optimized geometries is 2.0 kcal/mol; for stacked structures, it is 2.3 kcal/mol. The EFP2 energy components show that dispersion is a significant attractive force between nucleotide bases – even among the hydrogen bonded structures, in which the Coulomb energy is the single largest attractive energy term. For hydrogen bonded adenine-thymine and its methylated analog, the magnitude of the Coulomb term exceeds that of the exchange repulsion by only 1 to 3 kcal/mol. While the magnitude of the Coulomb term is about 9 kcal/mol higher than the exchange-repulsion term for the guanine-cytosine hydrogen bonded structure, this accounts for only about 30% of the total interaction energy, the rest being comprised of the dispersion and polarization terms. The stacked guanine-cytosine structure has a larger total interaction energy than the hydrogen bonded adenine-thymine structure.

An equation for the dispersion energy arising from the interaction between a molecule modeled with EFP and one modeled with a fully quantum mechanical (*ab initio*, AI) method was derived. The dynamic polarizability approach used for calculating the dispersion energy between two molecules modeled with EFP is too time-consuming to be used in the EFP-AI case. Effective distributed C_6^{ij} coefficients, between localized molecular orbitals *i* on the ab initio molecule and *j* on the EFP fragment, can be produced; these coefficients agree well with EFP-EFP distributed C_6 values. A total dispersion energy could be calculated by multiplying each C_6^{ij} by R_{ij}^{-6}, where R_{ij} is the distance between the centroids (centers of charge) of LMOs *i* and *j*.

ACKNOWLEDGMENTS

Many thanks to my advisor, Dr. Mark Gordon, and to past and present members of my graduate committee, including Dr. Theresa Windus, Dr. William Jenks, Dr. Jim Evans, Dr. Pat Thiel, Dr. Hans Stauffer, and Dr. Ricky Kendall. I am especially grateful to Dr. Lyudmila Slipchenko, to whom I owe the better part of my Ph.D., and to the ineffable Dr. Klaus Reudenberg, to whom I owe the rest. I also gratefully acknowledge Dr. Bosiljka Njegic, Dr. Federico Zahariev, and Dr. Albert DeFusco, along with the rest of my fellow Gordon Group members. Thanks also to the administrative staff at Iowa State University for their continued assistance in navigating the bureaucracy. As I could not have advanced this far in graduate school without building on the foundation formed during my undergraduate days, I would like to thank those who helped construct that foundation: Dr. Jens Nielsen, Dr. Nathan Baker, Dr. J. Andrew McCammon, and Kyle Wright. Finally, many thanks to my friends and family for their love and support.